UNITED STATES PATENT AND TRADEMARK OFFICE
_____

BEFORE THE PATENT TRIAL AND APPEAL BOARD
_____

EMC CORPORATION,
Petitioner,

v.

PERSONALWEB TECHNOLOGIES, LLC and
LEVEL 3 COMMUNICATIONS, LLC,
Patent Owners.

_____

Case IPR2013-00085
Patent 7,945,539 B2

_____

Before KEVIN F. TURNER, JONI Y. CHANG, and
MICHAEL R. ZECHER, *Administrative Patent Judges*.

CHANG, *Administrative Patent Judge*.

FINAL WRITTEN DECISION
35 U.S.C. § 318(a) and 37 C.F.R. § 42.73

I.   INTRODUCTION

EMC Corporation ("EMC") filed a petition on December 16, 2012, requesting an *inter partes* review of claims 10, 21, and 34 of U.S. Patent No. 7,945,539 B2 ("the '539 patent"). Paper 5 ("Pet."). PersonalWeb Technologies, LLC and Level 3 Communications, LLC (collectively, "PersonalWeb") filed a patent owner preliminary response. Paper 11 ("Prelim. Resp."). Taking into account the patent owner preliminary response, the Board determined that the information presented in the petition demonstrated that there was a reasonable likelihood that EMC would prevail with respect to at least one claim. Pursuant to 35 U.S.C. § 314, the Board instituted this trial as to claims 10, 21, and 34 of the '539 patent. Paper 18 ("Dec.").

After institution, PersonalWeb filed a patent owner response (Paper 40 ("PO Resp.")), and EMC filed a reply to the patent owner response (Paper 48 ("Reply")). Oral hearing was held on December 16, 2013.[1]

We have jurisdiction under 35 U.S.C. § 6(c). This final written decision is entered pursuant to 35 U.S.C. § 318(a). We hold that claims 10, 21, and 34 of the '539 patent are unpatentable under 35 U.S.C. §§ 102 and 103.

---

[1] This proceeding, as well as IPR2013-00082, IPR2013-00083, IPR2013-00084, IPR2013-00086, and IPR2013-00087, involve the same parties and similar issues. The oral arguments for all six *inter partes* reviews were merged and conducted at the same time. A transcript of the oral hearing is included in the record as Paper 72.

A. Related Proceeding

EMC indicates that the '539 patent is the subject of litigation titled *PersonalWeb Technologies LLC v. EMC Corporation and VMware, Inc.*, No. 6:11-cv-00660-LED (E.D. Tex.). Pet. 1.

B. The '539 patent

The '539 patent relates to a method for identifying a data item (e.g., a data file or record) in a data processing system, by using an identifier that depends on all of the data in the data item and only on the data in the data item. Ex. 1001, 1:45-48; 3:52-56. Thus, the identity of a data item is said to be independent of its name, origin, location, and address. *Id*. at 3:55-58. According to the '539 patent, the system provides transparent access to any data item by reference only to its identity and independent of its present location. *Id*. at 4:11-13. Figure 10(b) of the '539 patent, reproduced below, is a flow chart for determining an identifier of a data item.

FIG. 10(b)
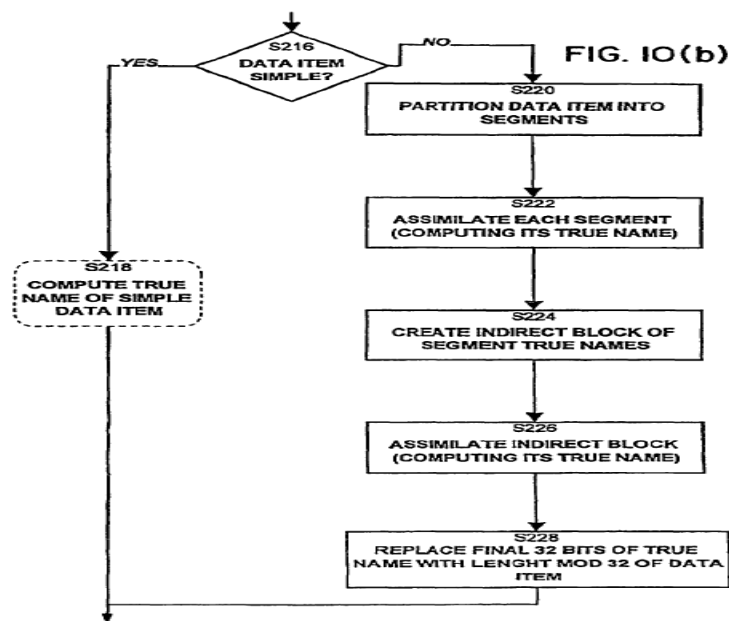
As shown in Figure 10(b) of the '539 patent, for a simple data item (a data item whose size is less than a particular given size) (S216 and S218), a data identifier (True Name) is computed using a function (e.g., a message digest ("MD") function, such as MD4 or MD5, or a secure hash algorithm ("SHA") function). *Id*. at 14:24-50, 15:37-48, figs. 10(a) & 10(b). As a result, a data item that has an arbitrary length is reduced to a relatively small, fixed size identifier (True Name) that represents the data item. *Id*.

If the data item is a compound data item (a data item whose size is greater than the particular given size), the system will partition the data item into segments (S220); assimilate each segment (S222); compute the True Name of the segment; create an indirect block consisting of the computed segment True Names (S224); assimilate the indirect block (S226); and replace the final 32-bits of the resulting True Name by the length modulo 32 of the compound data item (S228). *Id*. at 15:49-67, fig. 10(b). The result is the True Name of the compound data item. *Id*.

Figure 11 of the '539 patent is reproduced below:

FIG. 11

S230 DETERMINE TRUE NAME

S232 DOES TRUE NAME EXIST IN TRUE FILE REGISTRY?

S236
* CREATE NEW ENTRY
* SET USE COUNT TO 1
* STORE FILE ID
* SET OTHER FIELDS

S237 DOES ENTRY HAVE FILE ID?

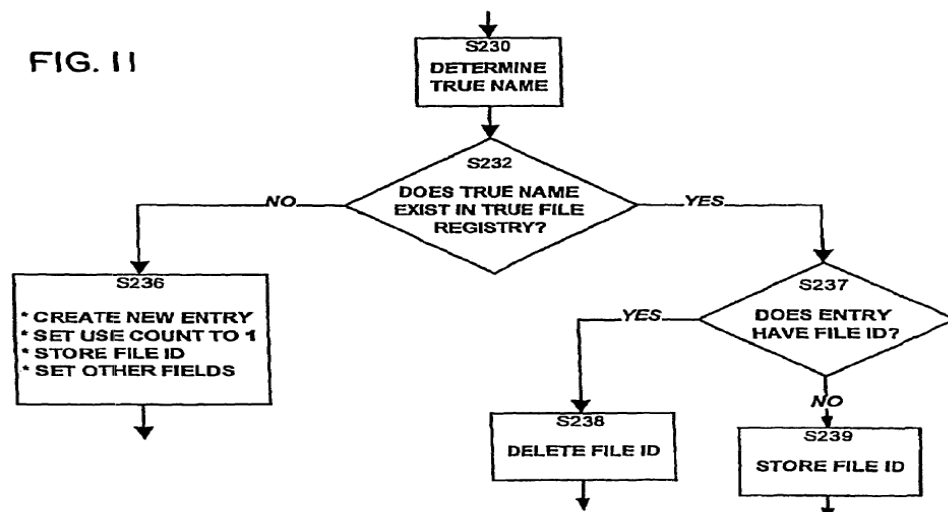S238 DELETE FILE ID

S239 STORE FILE ID

Figure 11 of the '539 patent depicts a mechanism for assimilating a data item into a file system. The purpose of this mechanism is to add a given data item to the True File registry. *Id*. at 16:10-16. If the data item already exists in the registry, the duplicate will be eliminated. *Id*.

To assimilate a data item, the system will determine the True Name of the data item corresponding to the file (S230); look for an entry for the True Name in the True File Registry (S232); and determine whether a True Name entry exists in the True File Registry (S232). *Id*. at 16:10-29, fig. 11. If the entry record includes a corresponding True File ID (Step S237), the system will delete the file (Step S238). *Id*. Otherwise, the system will store the True File ID in the entry record (S239). *Id*. If there is no entry in the True File Registry for the True Name (S232), the system will create a new entry in the True File Registry for the True Name (S236). *Id*.

C. Illustrative Claim

All of the challenged claims are independent claims. Claim 21 is illustrative and reproduced as follows:

> 21. A computer-implemented method of obtaining access to a data item at a first computer in a network of computers, said data item comprising a plurality of segments, each of said plurality of segments being stored on at least one of a plurality of computers in said network, said plurality of computers being distinct from said first computer, the method comprising the steps of:
>
> (A) by hardware in combination with software, using a first data identifier to obtain a plurality of segment identifiers, each of said segment identifiers corresponding to one of said plurality of segments, the segment identifier for each particular

5

segment being based at least in part on a first given function of the data comprising said particular segment and only the data in said particular segment, where any two identical segments will have identical segment identifier as determined using said first given function, and

wherein said first data identifier is based, at least in part, on a second given function of data comprising the plurality of segment identifiers;

(B) using the plurality of segment identifiers obtained in step (A) to obtain at least one of said plurality of segments by, for at least one particular segment identifier of said plurality of segment identifiers:

(b0) using said particular segment identifier to ascertain one or more locations in said network of computers that should have the corresponding particular segment;

(bl) using said particular segment identifier to request said corresponding particular segment from at least one of said one or more locations ascertained in step (b0); and

(b2) obtaining said corresponding particular segment from at least one location in said network.

Ex. 1001, 42:52-43:18.

D. Prior Art Relied Upon

EMC relies upon the following prior art references:

Woodhill	US 5,649,196[2]	July 15, 1997	(Ex. 1005)
Fischer	US 5,475,826[3]	Dec. 12, 1995	(Ex. 1036)

---

[2] Woodhill claims the benefit of U.S. patent application No. 08/085,596, filed on July 1, 1993.

[3] Fischer was filed on Nov. 19, 1993.

Albert Langer, "*Re: dl/describe (File descriptions)*," posted to the "alt.sources" "comp.archives.admin" newsgroups on Aug. 7, 1991 ("Langer," Ex. 1003)

Frederick W. Kantor, "*FWKCS (TM) Contents_Signature System Version 1.22*," FWKCS122.REF (Aug. 10, 1993) ("Kantor," Ex. 1004)

E. Grounds of Unpatentability

The Board instituted the instant trial based on the following grounds of unpatentability:

Claim	Basis	References
10 and 21	§ 102(b)	Langer
34	§ 103(a)	Langer and Woodhill
10 and 21	§ 103(a)	Kantor
34	§ 103(a)	Kantor and Langer
10 and 21	§ 103(a)	Woodhill and Fischer

II. ANALYSIS

A. Claim Construction

We begin our analysis by determining the meaning of the claims. In an *inter partes* review, claim terms in an unexpired patent are given their broadest reasonable construction in light of the specification of the patent in which they appear. 37 C.F.R. § 42.100(b). Under the broadest reasonable construction standard, claim terms are given their ordinary and customary meaning as would be understood by one of ordinary skill in the art in the context of the entire disclosure. *In re Translogic Tech. Inc.*, 504 F.3d 1249, 1257 (Fed. Cir. 2007).

The parties proposed a claim construction for each of the following claim terms: (1) "data" and "data item," (2) "data identifier," (3) "True Name," and (4) "location." Pet. 5-6; Prelim. Resp. 3-5. In the Decision on Institution, we addressed the parties' proposed claim constructions and set forth the broadest reasonable interpretation of each of the claim terms. Dec. 8-13. Neither party challenges our claim constructions. PO Resp. 1-2; Reply in general. We discern no reason to deviate from those constructions for the purposes of this decision. For convenience, the claim constructions proffered in the Decision on Institution are set forth in the table below.

Claim Terms	Claim Constructions
data item	Sequence of bits which includes one of the following: (1) the contents of a file; (2) a portion of a file; (3) a page in memory; (4) an object in an object-oriented program; (5) a digital message; (6) a digital scanned image; (7) a part of a video or audio signal; (8) a directory; (9) a record in a database; (10) a location in memory or on a physical device or the like; and (11) any other entity which can be represented by a sequence of bits. Dec. 8-10.
data	A subset of a data item. *Id.* at 9-10.
data identifier	A substantially unique alphanumeric label for a particular data item. *Id.* at 10-11.
True Name	A substantially unique alphanumeric label for a particular data item. *Id.* at 12-13.
location	Any of a particular processor in the system, a memory of a particular process, a storage device, a removable storage medium (such as a floppy disk or compact disk), or any physical location in the system. *Id.* at 13.

After institution, PersonalWeb asserts that the preambles of claims 10 and 21 are limiting. PO Resp. 2-3. PersonalWeb argues that each claim body refers back to its preamble for completeness. *Id*. We agree. In general, a preamble is construed as a limitation "if it recites essential structure or steps, or if it is 'necessary to give life, meaning, and vitality' to the claim." *Catalina Mktg. Int'l, Inc. v. Coolsavings.com, Inc.,* 289 F.3d 801, 808 (Fed. Cir. 2002) (quoting *Pitney Bowes, Inc. v. Hewlett-Packard Co.,* 182 F.3d 1298, 1305 (Fed. Cir. 1999)). When the limitations in the body of the claim "rely upon and derive antecedent basis from the preamble, then the preamble may act as a necessary component of the claimed invention." *Eaton Corp. v. Rockwell Int'l Corp.*, 323 F.3d 1332, 1339 (Fed. Cir. 2003).

Notably, the preamble of claim 10 recites "[a] computer-implemented method of obtaining access to *a data item* at a first computer in *a network of computers*, said data item comprising *a plurality of segments*." Ex. 1001, 41:57-60. The preamble of independent claim 21 recites similar features. Indeed, the body of claim 10, which includes the claim terms "said data item," "said network of computers," and "said plurality of segments," relies upon and derives antecedent basis from the preamble of claim 10. Similarly, the body of claim 21, which includes the claim terms "said network of computers" and "said plurality of segment," relies upon and derives antecedent basis from the preamble of claim 21.

Accordingly, we conclude that the preambles of claims 10 and 21 are entitled to patentable weight.

B. Whether Kantor and Langer are "Printed Publications"

In its petition, EMC takes the position that Kantor and Langer each are a "printed publication" within the meaning of 35 U.S.C. § 102(b). Pet. 35-36, 42. EMC asserts that Kantor has been publicly available since August 1993, which is prior to the critical date, April 11, 1995, one year before the earliest priority date claimed by the '539 patent. *Id*. at 4, n.3. EMC also submits that Langer has been publicly available before the critical date, because Langer was made available on the "alt.sources.d" and "comp.archives.admin" Usenet newsgroups on August 7, 1991. *Id*. at 3, n.2, 36. As support, EMC proffers declarations of Mr. Michael A. Sussell (Ex. 1053), Mr. Jason S. Sadofsky (Exs. 1081, 1091), and Mr. Keith Moore (Ex. 1059) to confirm the publication and authenticity of Kantor and Langer.

PersonalWeb counters that neither Kantor nor Langer is a "printed publication." PO Resp. 54-60. In particular, PersonalWeb alleges that EMC has not established that the references existed prior to the critical date, because EMC's witnesses did not review the references before the critical date. *Id*. at 55-57. PersonalWeb also contends that there is no evidence that the references were disseminated publicly, catalogued, or indexed in a meaningful way. *Id*. PersonalWeb maintains that EMC fails to establish that one with ordinary skill in the art, exercising reasonable diligence, would have located the documents prior to the critical date. *Id*.

Based on the evidence before us, we are not persuaded by PersonalWeb's arguments. Rather, we determine that EMC has demonstrated sufficiently that Kantor and Langer are "printed publications."

The determination of whether a given reference qualifies as a prior art "printed publication" involves a case-by-case inquiry into the facts and circumstances surrounding the reference's disclosure to members of the public. *In re Klopfenstein*, 380 F.3d 1345, 1350 (Fed. Cir. 2004). The key inquiry is whether the reference was made "sufficiently accessible to the public interested in the art" before the critical date. *In re Cronyn*, 890 F.2d 1158, 1160 (Fed. Cir. 1989); *In re Wyer*, 655 F.2d 221, 226 (CCPA 1981). "A given reference is 'publicly accessible' upon a satisfactory showing that such document has been disseminated or otherwise made available to the extent that persons interested and ordinarily skilled in the subject matter or art exercising reasonable diligence, can locate it . . . ." *Bruckelmyer v. Ground Heaters, Inc.*, 445 F.3d 1374, 1378 (Fed. Cir. 2006) (citation omitted).

Indexing is not "a necessary condition for a reference to be publicly accessible," but is only one among many factors that may bear on public accessibility. *In re Lister*, 583 F.3d 1307, 1312 (Fed. Cir. 2009). In that regard, "while often relevant to public accessibility, evidence of indexing is not an absolute prerequisite to establishing online references . . . as printed publications within the prior art." *Voter Verified, Inc. v. Premier Election Solutions, Inc.,* 698 F.3d 1374, 1380 (Fed. Cir. 2012).

Contrary to PersonalWeb's assertion that Kantor did not exist prior to the critical date and there is no evidence that Kantor was disseminated publicly, Kantor itself shows a copyright date of "1988-1993" and a posted date of "1993 August 10." Ex. 1004, Title Page, the first page after the Title

Page ("All of the programs and documents, comprising the entire contents of this Authenticity Verification Zipfile FWKCS122.ZIP, together with this Zipfile itself, are, in accordance with their respective dates of creation or revision, (C) Copyright Frederick W. Kantor 1988-1993."). Kantor also states:

> The FWKCS(TM) Contents_Signature System has become a robust platform for supporting contents_signature functions. FWKCS provides many functions and options for application in a public, commercial, school, institutional, or governmental environment. Extensive technical support is of special value in helping such users to benefit more fully from these many features.

> Registered FWKCS hobby BBS users are able to receive a modest amount of assistance, and are invited to participate in the FWKCS conference on The Invention Factory BBS, echoed via Execnet.

> Commercial, school, institutional, and governmental users, with their special support needs, are invited to discuss terms for obtaining such assistance.

> . . . .

> To get a new version of FWKCS, download FWKCSnnn.ZIP from The Invention Factory BBS, where nnn is the new version number without a decimal point. These special downloads are available at no fee, from a 43_line hunt_up group of USR Dual Standard modems, at 2400-16800 bits/sec (including V32.bis).

Ex. 1004, 158-59. It is clear from Kantor that, during the 1988-1993 timeframe, Dr. Kantor had posted many versions of his software and user manual—including Kantor (version 1.22), the version relied upon by EMC (Ex. 1004)—on electronic Bulletin Board Systems.

Mr. Sussell, the co-owner and system operator of the Invention Factory Bulletin Board System, testifies that the Invention Factory Bulletin Board System is a computer system that allows users to share files, messages, and articles, as well as search, upload, and download files. Ex. 1053 ¶¶ 3, 4. According to Mr. Sussell, he and his wife launched the Invention Factory Bulletin Board System in 1983, and it had over 3,000 subscribers by mid-1993. *Id.* at ¶ 6. Mr. Sussell testifies that, by 1993, the system provided all users keyword search functionality and access to various descriptive and meaningful directories. *Id.* at ¶¶ 8-10.

Importantly, Mr. Sussell testifies that the Invention Factory Bulletin Board System "extensively utilized and hosted current versions of FWKCS software on its [Bulletin Board System]," and "made publicly accessible and available the complete FWKCS ZIP file that contained both the software as well as related documentation such as user manuals" prior to the critical date. *Id.* at ¶ 15; *see also id.* at ¶¶ 16-27. Specifically, Mr. Sussell testifies that users would have found Kantor by performing keyword searches on the Invention Factory Bulletin Board System. *Id.* at ¶ 21. Mr. Sussell also indicates that the Invention Factory Bulletin Board System advertised Dr. Kantor's software to its users by including information about the software on the "Welcome" screen, and made the FWKCS Zip file available in four different directories. *Id.* at ¶¶ 18-20. Mr. Sussell further testifies that computer disks that contain the FWKCS Zip file were distributed at various Bulletin Board System conferences. *Id.* at ¶ 18.

13

Mr. Sadofsky, a technology archivist and software historian, testifies
that he personally verified the authenticity of Kantor—the user manual
(version 1.22), the version relied upon by EMC (Ex. 1004)—by comparing it
with a "1993 archived" version, and determined that Kantor is identical to
the "1993 archived" version. Ex. 1081 ¶¶ 14-17. Mr. Sadofsky testifies that
the source file of the "1993 archived" version has a timestamp of
August 10, 1993, at 1:22 AM. *Id.* at ¶ 16; *see also* Ex. 1091 ¶¶ 10-11;
Ex. 2014 ¶ 5. According to Mr. Sadofsky, Kantor was publicly accessible
prior to the critical date. Ex. 1081 ¶¶ 13, 16-17.

PersonalWeb also asserts that Kantor was buried and hidden in the zip
file in a manner such that "it would not have been located and accessed by
persons interested and ordinarily skilled in the art exercising reasonable
diligence even if they had access to the ZIP file." PO Resp. 58-59 (citing
Ex. 2014). However, PersonalWeb's supporting evidence, Mr. Thompson's
declaration (Ex. 2013, 2014), does not substantiate PersonalWeb's assertion.
Upon review of Mr. Thompson's declaration, we observe that
Mr. Thompson downloaded the FWKCS Zip file without any difficultly.
Ex. 2014 ¶ 5. Significantly, Mr. Thompson did not follow the instructions
provided with the zip file, nor did he use the appropriate computer
environment (DOS 3.0 or an IBM OS/2 2.0) that was used normally in
1993-1994 timeframe. Ex. 2014 ¶¶ 6-11; Ex. 1091 ¶¶ 5, 14. Instead, he
used non-compatible software (DOS 8.0 and 32-bit Windows XP operating
system that was released in 2001). *Id.* Once he followed the instructions

and unzipped the FWKCS Zip file, Mr. Thompson located Kantor without difficulty. Ex. 2014 ¶¶ 20-22.

Mr. Sadofsky confirms that the README.TXT file provides simple instructions and, if a user follows the instructions and uses the operating system that was used normally in 1993-1994 timeframe, the user could locate Kantor without difficulty. Ex. 1091 ¶¶ 13-17. In fact, Mr. Sadofsky demonstrated, in his declaration, several relatively easy ways for a user to access Kantor—with or without installing the software, and with or without help screens. Ex. 1091 ¶¶ 8-16 (II. README.TXT); ¶¶ 17-20 (III. GETLOOK.BAT); ¶¶ 21-22 (IV. FWKCS122 Start Screen and In-Program Help). Based on the evidence before us, we determine that Kantor was available to the extent that persons interested and ordinarily skilled in the art, exercising reasonable diligence, could locate it.

The evidence on this record sufficiently supports that Kantor was posted on a publicly accessible site—the Invention Factory Bulletin Board System—well known to those interested in the art and could be downloaded and retrieved from that site, and, therefore, Kantor, an electronic publication, is considered a "printed publication" within the meaning of 35 U.S.C. § 102(b). *See Wyer*, 655 F.2d at 226 (An electronic publication, including an on-line database or Internet publication, is considered to be a "printed publication" "upon a satisfactory showing that such document has been disseminated or otherwise made available to the extent that persons interested and ordinarily skilled in the subject matter or art, exercising reasonable diligence, can locate it and recognize and comprehend therefrom

the essentials of the claimed invention without need of further research or experimentation.").

Similarly, the evidence on this record shows Langer was publicly distributed prior to the critical date. The header on the first page of Langer, reproduced below, indicates that Langer was distributed on August 7, 1991 to the newsgroups "alt.sources.d" and "comp.archives.admin" (Ex. 1003, 1):

```
From: cmf851@anu.oz.au (Albert Langer)
Newsgroups: alt.sources.d,comp.archives.admin
Subject: Re: dl/describe (File descriptions) posted to alt.sources
Message-ID: <1991Aug7.225159.786@newshost.anu.edu.au>
Date: 7 Aug 91 22:51:59 GMT
References: <1991Aug7.124457.6814@csv.viccol.edu.au>
<1991Aug7.131048.6817@csv.viccol.edu.au>
Sender: news@newshost.anu.edu.au
Followup-To: comp.archives.admin
Organization: Computer Services Centre, Australian National University,
Canberra, Australia.
Lines: 291
```

Mr. Moore, who has personal knowledge of the operation of Usenet in 1991, testifies that Langer's header is consistent with the format of Usenet articles from the 1991 time frame, and the "Date:" field—indicating that Langer was posted on August 7, 1991, at approximately 10:51 PM GMT—would have generated automatically when the article was posted to Usenet. Ex. 1059 ¶ 16. Mr. Moore also testifies that he personally verified the authenticity of Langer by comparing it with an archived version obtained from Google Groups. *Id*. at ¶ 19.

According to Mr. Moore, Usenet was a network of computers that individuals could use to send and receive technical articles. *Id*. at ¶ 13. In particular, Mr. Moore indicates that anyone could subscribe to a Usenet

newsgroup without restrictions, and that subscribers could read articles from the Usenet newsgroups. *Id*. at ¶ 21. Mr. Moore testifies that Langer was distributed through two specific Usenet newsgroups: (1) "alt.sources.d," which hosted technical discussions about source code; and (2) "comp.archives.admin," which focused on technical issues related to the administration of computer archives. *Id*. at ¶ 22. Mr. Moore further declares that Usenet articles were distributed automatically to the registered readers, and, during the 1991-1992 timeframe, the "alt.sources.d" and "comp.archives.admin" newsgroups had 37,000 and 27,000 registered readers, respectively. *Id*. at ¶¶ 23-24. Given the evidence before us, we determine that EMC has established sufficiently that Langer was distributed publicly to those interested in the art.

We also are not persuaded by PersonalWeb's argument that EMC's witnesses personally did not post or review Kantor and Langer prior to the critical date. PO Resp. 55-57 (citing Ex. 2015, 52-55; Ex. 2013, 29-30; Ex. 2016, 98, 180; Ex. 2019, 49-50). It is well settled that it is not necessary for the witnesses to have reviewed the reference personally prior to the critical date in order to establish publication. *See In re Hall*, 781 F.2d 897, 899 (Fed. Cir. 1986) (concluding "that competent evidence of the general library practice may be relied upon to establish an approximate time when a thesis became accessible"); *Wyer*, 655 F.2d at 226 (Notwithstanding that there is no evidence concerning actual viewing or dissemination of any copy of the Australian application, the court held that "the contents of the application were sufficiently accessible to the public and to persons skilled

17

in the pertinent art to qualify as a 'printed publication.'"); *In re Bayer*, 568 F.2d 1357, 1361 (CCPA 1978) (A reference constitutes a "printed publication" under 35 U.S.C. § 102(b) as long as a presumption is raised that the portion of the public concerned with the art would know of the invention.).

For the foregoing reasons, we determine that EMC has demonstrated, by a preponderance of the evidence, that Kantor and Langer are "printed publications" within the meaning of 35 U.S.C. § 102(b). Therefore, EMC may rely upon Kantor and Langer for its asserted grounds of unpatentability under 35 U.S.C. §§ 102(b) and 103(a).

C. Principles of Law

To establish anticipation, each and every element in a claim, arranged as recited in the claim, must be found in a single prior art reference. *Net MoneyIN, Inc. v. VeriSign, Inc*., 545 F.3d 1359, 1369 (Fed. Cir. 2008); *Karsten Mfg. Corp. v. Cleveland Golf Co.*, 242 F.3d 1376, 1383 (Fed. Cir. 2001). We also recognize that prior art references must be "considered together with the knowledge of one of ordinary skill in the pertinent art." *In re Paulsen*, 30 F.3d 1475, 1480 (Fed. Cir. 1994) (citation and internal quotation marks omitted). Moreover, "it is proper to take into account not only specific teachings of the reference but also the inferences which one skilled in the art would reasonably be expected to draw therefrom." *In re Preda*, 401 F.2d 825, 826 (CCPA 1968).

A patent claim is unpatentable under 35 U.S.C. § 103(a) if the differences between the claimed subject matter and the prior art are such that the subject matter, as a whole, would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 406 (2007). The question of obviousness is resolved on the basis of underlying factual determinations, including: (1) the scope and content of the prior art; (2) any differences between the claimed subject matter and the prior art; (3) the level of skill in the art; and (4) secondary considerations of nonobviousness. *Graham v. John Deere Co.*, 383 U.S. 1, 17-18 (1966). The level of ordinary skill in the art is reflected by the prior art of record. *See Okajima v. Bourdeau*, 261 F.3d 1350, 1355 (Fed. Cir. 2001); *In re GPAC Inc.*, 57 F.3d 1573, 1579 (Fed. Cir. 1995); *In re Oelrich*, 579 F.2d 86, 91 (CCPA 1978).

We analyze the instituted grounds of unpatentability in accordance with the above-stated principles.

D. Claims 10 and 21 – Anticipated by Langer

EMC asserts that claims 10 and 21 are unpatentable under 35 U.S.C. § 102(b) as anticipated by Langer. Pet. 35-40. As support, EMC provides detailed explanations as to how each claim element, arranged as recited in the claim, is disclosed by Langer. *Id.* EMC also relies upon the declaration of Dr. Douglas W. Clark. Ex. 1009 ¶¶ 26-29.

Langer

Langer discloses a method of accessing files in a network of computers. Ex. 1003, 3. For instance, a file request may be embedded in a news article and include a unique identifier for the file. *Id*. at 3-4. As a result, users are informed automatically about the nearest location of the file. *Id*. Langer further discloses that a unique identifier for a file is calculated using a hash function (e.g., MD5, a cryptographic hash function) on the entire contents of the file, rather than the file's location. *Id*. at 2-3. For a package (e.g., an archive, which is a collection of files packaged together) that is divided into its component files, a unique identifier for each component file is calculated by using an MD5 hash function on the contents of the component file. *Id*. at 5. The unique identifier for the entire package is calculated by applying *an MD5 hash again to the concatenation of the MD5 hashes* of the component files ("a hash of hashes"). *Id*.

Discussion

In its patent owner response, PersonalWeb counters that Langer does not describe obtaining a plurality of segment identifiers, or a segment, in response to a request comprising the first identifier, as required by claims 10 and 21. PO Resp. 40-42. As support, PersonalWeb proffers a declaration of Dr. Robert B. K. Dewar. Ex. 2020 ¶¶ 75-78.

Claim 10 recites "in response to a request, said request comprising a first identifier, obtaining a plurality of segment identifiers, . . . using at least one of said segment identifiers . . . , requesting at least one particular

20

segment of said plurality of segments that comprise said data item."
Ex. 1001, 41:61-42:10. Claim 21 recites similar limitations.

In its petition, EMC takes the position that Langer meets the limitation because Langer describes a method that, in response to a request including the MD5 hash code of the package, obtains the MD5 hash codes for the inner files of the package. Pet. 39 (citing Ex. 1009 ¶ 29; Ex. 1003, 3-5). As support, Dr. Clark testifies that a person with ordinary skill in the art would have understood that, in order to retrieve a particular inner file of the package, the MD5 hash code of the entire package could be used to obtain the MD5 hash codes that were computed for the inner files. Ex. 1009 ¶ 29 (citing Ex. 1003, 4).

PersonalWeb does not disagree that: (1) Langer's package is a data item; (2) the individual inner files of a package are segments; (3) the MD5 codes of the inner files are the segment identifiers; and (4) an MD5 code of the concatenation of the codes of the inner files from the package is the first identifier. PO Resp. 40 (citing Pet. 39; Ex. 1046, 3-6).

However, PersonalWeb alleges that Langer fails to disclose:
(1) accessing an inner file of a package by sending a request that includes an MD5 code of the package; (2) obtaining a plurality of MD5 codes of the inner files in response to such a request; and (3) using one of MD5 codes to obtain a particular inner file of the package. *Id*. at 41 (citing Ex. 2020 ¶¶ 75, 76). PersonalWeb contends that EMC improperly attempts to switch back and forth between Langer's package embodiment (Ex. 1003, 5-6) and Langer's standalone file embodiment (Ex. 1003, 3-4). PO Resp. 42. In

particular, PersonalWeb maintains that Langer's package embodiment does not meet the disputed limitation because the technique of using MD5 codes is not used to access an inner file of a package, and that Langer's standalone file embodiment also does not meet the disputed limitation because it does not have a plurality of segment identifiers. *Id*. at 42 (Ex. 2020 ¶¶ 76, 77).

In its reply, EMC maintains that Langer explicitly discloses that a user may submit a query to a database using an MD5 code to determine the location of a file so it could be retrieved. Reply 8 (citing Ex. 1003, 3-4). EMC asserts that the MD5 code of a package may be used to obtain the concatenated block of MD5 codes, or a listing of MD5 codes and filenames of the inner files of the package. *Id*. (citing Ex. 1003, 5-6; Ex. 1092 ¶¶ 49, 50; Ex. 1088, 381-82). According to EMC, in either situation, an MD5 code for a particular inner file of the package may then be used to identify and retrieve that particular inner file. *Id*. (citing Ex. 1003, 3-4; Ex. 1009 ¶¶ 28, 29). We agree with EMC.

We observe that PersonalWeb's arguments and expert testimony essentially rest on the incorrect premise that Langer has two separate and distinct embodiments. We disagree with PersonalWeb's assertion that Langer's disclosure under the heading of "Unique Identifiers" (Ex. 1003, 3-4) is a standalone file embodiment, and has little to do with other portions of Langer. In fact, that disclosure of Langer merely teaches the overall concept of utilizing unique identifiers (e.g., MD5 hash codes) to access files. Nothing in Langer indicates that the unique identifiers (Ex. 1003, 3-4) are limited to standalone files, and could not apply to files within a package.

Moreover, we are not persuaded by PersonalWeb's argument that Langer teaches away from the disputed limitation (PO Resp. 41-42). At the outset, we note that, although a "teaching away" argument could be relevant to an obviousness analysis, "whether a reference teaches away from an invention is inapplicable to an anticipation analysis." *ClearValue, Inc. v. Pearl River Polymers, Inc.*, 668 F.3d 1340, 1344 (Fed. Cir. 2012) (quoting *Celeritas Techs., Ltd. v. Rockwell Int'l Corp.*, 150 F.3d 1354, 1361 (Fed. Cir. 1998)) (internal quotation marks omitted).

In any event, PersonalWeb's argument and expert testimony contradict Langer's explicit disclosure. Notably, Dr. Dewar testifies that a person with ordinary skill in the art would have recognized that the MD5 codes of the inner files "would have been *calculated locally* upon receipt of a new package in order to allow a user to see if the codes for a new package matched those of another package – not for accessing files of a package." Ex. 2020 ¶ 76 (emphasis added). As noted by EMC, however, such local calculation would require *downloading the entire package*, contrary to Langer's stated objective for his invention. Reply 9. Indeed, Langer discloses that the files are distributed among different locations, and acknowledges the inefficiency of obtaining the entire package when a user has a new MD5 code for a package. Ex. 1003, 5 ("[I]t would be nice to be able to tell the user *without the need for collecting the entire package*." (emphasis added)). Therefore, in the absence of an explicit disclosure, one with ordinary skill in the art would not have read Langer to download the

entire package and calculates the MD5 codes *locally*, as PersonalWeb alleges.

On the other hand, we are persuaded by EMC's contention that an MD5 code for a particular file within a package may be used to identify and retrieve that particular file. As Dr. Clark testifies, Langer discloses that, in response to a request for a file from a user, the central database server uses the MD5 code to return the locations that store a copy of the file corresponding to the MD5 code. Ex. 1009 ¶ 29. Dr. Clark's testimony is consistent with the express disclosure of Langer, which provides that "[a]n archie or similar lookup could first determine which nearby systems have the file," and "that database lookup may as well also provide the local directory and filename for it." Ex. 1003, 4.

Langer also explains that a package's unique identifier is computed by hashing the MD5 codes for the individual files within the package. Ex. 1003, 5. Langer describes a technique that eliminates the need for the user to download the entire package even in the situation where a package has a new MD5 code, and an individual file in a package can be requested. *Id*. at 5-6. Langer further describes that the central database (e.g., archie) explodes the contents of package files and lists the individual items within them. *Id*. at 5. According to Dr. Clark, the MD5 code of the package is used to obtain a listing of MD5 codes and filenames of the inner files in the package. Ex. 1092 ¶ 50 (citing Ex. 1088, 381-82). We credit Dr. Clark's testimony in that regard because it is consistent with Langer's disclosure. Ex. 1003, 5. Therefore, we agree with EMC that an MD5 code for a

particular file within the package may then be used to identify and retrieve that particular file. Reply 8 (citing Ex. 1003, 3-4; Ex. 1009 ¶¶ 28, 29).

For the foregoing reasons, we determine that EMC has demonstrated by a preponderance of the evidence that claims 10 and 21 are anticipated by Langer.

E. Claim 34 – Obvious Over Langer and Woodhill

EMC asserts that claim 34 is unpatentable under 35 U.S.C. § 103(a) as obvious over Langer and Woodhill. Pet. 35-41 (citing Ex. 1003, 3-5; Ex. 1005, 15:13-20). In support of that asserted ground of unpatentability, EMC provides explanations as to how each claim limitation is taught or suggested by the combination of Langer and Woodhill, and a rationale for combining the references. *Id*. EMC also relies upon its explanations regarding the anticipation ground of unpatentability based on Langer, and Dr. Clark's testimony. *Id*.; Ex. 1009 ¶¶ 28-31.

In its patent owner response, PersonalWeb counters that the obviousness ground of unpatentability does not cure the deficiencies of Langer. PO Resp. 43. PersonalWeb essentially relies upon the same arguments presented with respect to the anticipation ground of unpatentability as to claims 10 and 21. *Id*. As discussed above, we have addressed those arguments and determined that they are unavailing. PersonalWeb further argues that the combination of Langer and Woodhill does not disclose all of the limitations of claim 34, and alleges that there is insufficient reason to combine Langer and Woodhill.

Based on the evidence before us, we are not persuaded by PersonalWeb's arguments. In the analysis below, we focus on the deficiencies alleged by PersonalWeb with respect to the ground of unpatentability based on the combination of Langer and Woodhill.

Dividing a data item into a plurality of segments

Claim 34 recites "dividing a particular data item into a plurality of segments; . . . determining a plurality of segment identifiers by . . . determining a corresponding segment identifier for each particular segment of said plurality of segments." Ex. 1001, 45:5-9.

In its petition, EMC maintains that dividing a file into a plurality of segments was a known, effective technique to handle large files, as evidenced by Woodhill, to reduce the amount of data that must be transmitted. Pet. 40-41 (citing Ex. 1009 ¶¶ 30-31; Ex. 1005, 15:13-20). According to EMC, one with ordinary skill in the art would have found it obvious to combine Langer's method of accessing files using unique identifiers (e.g., MD5 codes) and Woodhill's technique of dividing a file into a plurality of segments. *Id.*

In its patent owner response, PersonalWeb counters that Langer fails to disclose the aforementioned limitation recited in claim 34. PO Resp. 43-44. In particular, PersonalWeb argues that Langer does not determine segment identifiers for *all segments* of a package, because "no hash function or algorithm is applied to any directory, directory tree, or header of any package in Langer." *Id.* 44-45 (citing Ex. 2020 ¶ 84). PersonalWeb further alleges that Langer teaches away from the subject matter of claim 34, as

26

Langer specifically states that it intentionally does not apply a hash function to the directory. *Id*. at 46 (citing Ex. 1003, 5; Ex. 2020 ¶ 85).

We are not persuaded by PersonalWeb's arguments and supporting evidence. As EMC points out (Reply 4-6, 9-10), PersonalWeb's arguments are based on an unreasonable construction of the claim term "data item" that requires the data item to include directories and headers, which are not part of the *contents* of the inner files of a package. As discussed above, the broadest reasonable interpretation of the claim term "data item" is "a sequence of bit." We clarified in this decision and in the Decision on Institution (Dec. 8-10) that one of the examples of a data item is "a *portion* of a file." That example is consistent with the specification of the '539 patent. Ex. 1001, 2:16-21 ("[A] data item may be . . . a portion of a file."). Moreover, PersonalWeb consistently has agreed with that claim interpretation and example. Prelim. Resp. 3; PO Resp. 1-2. Therefore, claim 34 does not require a data item to include directories or headers of a package. Consequently, Langer's technique of determining unique identifiers by applying a hash function to the *contents* of the inner files, and not to directories or headers, describes the aforementioned limitation recited in claim 34.

Furthermore, Langer's technique does not criticize, discredit or otherwise discourage the aforementioned claim limitation. *See DePuy Spine, Inc. v. Medtronic Sofamor Danek, Inc.*, 567 F.3d 1314, 1327 (Fed. Cir. 2009) ("A reference does not teach away, however, if it merely expresses a general preference for an alternative invention but does not

'criticize, discredit, or otherwise discourage' investigation into the invention claimed.") (quoting *In re Fulton*, 391 F.3d 1195, 1201 (Fed. Cir. 2004)). Rather, we determine that Langer's technique meets the claim limitation, because the claim limitation does not require the data item or segments of the data item to include directories or headers.

Reasons to Combine Langer and Woodhill

In its patent owner response, PersonalWeb argues that it would not have been obvious to one with ordinary skill in the art to combine Woodhill and Langer. PO Resp. 47 (citing Ex. 2020 ¶ 87). In particular, PersonalWeb alleges that Woodhill's system is concerned with backing up files, but Langer never discloses any desire to back up files to a remote backup server. *Id*. PersonalWeb also contends that there is no reason to apply Woodhill's granularization technique that is related to large database files to Langer. *Id*.

We are not persuaded by PersonalWeb's arguments, as they improperly assume that Woodhill's disclosure is limited to backing up files and Langer's method is limited to small non-database files. Indeed, Langer does not place any requirement on the type or size of files. Ex. 1003, 3-5. The mere fact that the two references have different objectives does not mean that a person of ordinary skill in the art would not combine their teachings. *In re Heck*, 699 F.2d 1331, 1333 (Fed. Cir. 1983) ("The use of patents as references is not limited to what the patentees describe as their own inventions or to the problems with which they are concerned.") (quoting *In re Lemelson*, 397 F.2d 1006, 1009 (CCPA 1968)).

Importantly, a prior art reference must be considered for everything it teaches by way of technology and is not limited to the particular invention it is describing and attempting to protect. *EWP Corp. v. Reliance Universal Inc.*, 755 F.2d 898, 907 (Fed. Cir. 1985), *cert. denied*, 474 U.S. 843 (1985). EMC's proposed modification does not require incorporating Woodhill's *entire back-up procedure* into Langer's method. EMC merely relies upon Woodhill's *technique of dividing files* into a plurality of segments. *See KSR*, 550 U.S. at 420 ("[F]amiliar items may have obvious use beyond their primary purpose, and in many cases a person of ordinary skill will be able to fit the teachings of multiple patents together like pieces of a puzzle.").

Upon reviewing the record before us, we determine that EMC's suggestion for modifying Langer's method of accessing a file using an unique identifier with Woodhill's technique of dividing a file into a plurality of segments—to reduce the amount of data that must be transmitted and to provide a more efficient method of handling large data files—suffices as an articulated reason with a rational underpinning to justify the legal conclusion of obviousness. *See KSR*, 550 U.S. 416 ("The combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results.").

Secondary Considerations of Nonobviousness

PersonalWeb argues that its evidence of non-obviousness outweighs EMC's evidence of obviousness. PO Resp. 54. In support of its argument, PersonalWeb proffers three licensing agreements, as well as the declaration of Mr. Kevin Bermeister. *Id.* (citing Exs. 2010- 12; Ex. 2009 ¶¶ 3-9).

PersonalWeb argues that each license granted to a third party was not for the purpose of settling a patent infringement suit. *Id.*

In its Reply, EMC contends that PersonalWeb has failed to establish a sufficient nexus between the challenged claims of the '539 patent and the above-identified license agreements. Reply 14. EMC argues that each of the licenses granted rights to more than just the challenged claims, and involved related parties with interlocking ownership and business interests. *Id.* We agree with EMC that PersonalWeb has failed to establish the requisite nexus between the licensing agreements and the challenged claim.

A party relying on licensing activities as evidence of non-obviousness must demonstrate a nexus between those activities and the subject matter of the claims at issue. *GPAC*, 57 F.3d at 1580. Further, without a showing of nexus, "the mere existence of . . . licenses is insufficient to overcome the conclusion of obviousness" when there is a strong ground of unpatentability based on obviousness. *SIBIA Neurosciences, Inc. v. Cadus Pharm. Corp.*, 225 F.3d 1349, 1358 (Fed. Cir. 2000); *Iron Grip Barbell Co. v. USA Sports, Inc.*, 392 F.3d 1317, 1324 (Fed. Cir. 2004).

The evidence of non-obviousness presented by PersonalWeb falls short of demonstrating the required nexus. Neither PersonalWeb nor the declaration of Mr. Bermeister (Ex. 2009) establishes that the licensing agreements (Exs. 2010-12) are directed to the claimed subject matter recited in any of the challenged claims. For instance, PersonalWeb does not present credible or sufficient evidence that the three licensing agreements arose out of recognition and acceptance of the claimed subject matter recited in any of

the challenged claims. In the absence of an established nexus with the claimed invention, secondary consideration factors are entitled little weight, and generally have no bearing on the legal issue of obviousness. *See In re Vamco Machine & Tool, Inc.*, 752 F.2d 1564, 1577 (Fed. Cir. 1985). Furthermore, even if we assume that above-identified licenses establish some degree of industry respect for the claimed subject matter recited in challenged claims, that success is outweighed by the strong evidence of obviousness over the combination of Langer and Woodhill presented by EMC.

Based on this record, including the evidence of obviousness and the secondary considerations regarding licensing activities, we conclude that EMC has demonstrated by a preponderance of the evidence that claim 34 would have been obvious over the combination of Langer and Woodhill.

F. Claims 10 and 21 – Obvious Over Kantor

EMC asserts that claims 10 and 21 are unpatentable under 35 U.S.C. § 103(a) as obvious over Kantor. Pet. 42-49 (citing Ex. 1009 ¶¶ 31, 34-37, 39, 41-43; Ex. 1004, Preface, 7-11, 48, 51-55, 96-97, 173-74; Ex. 1047). As support, EMC provides explanations as to how each claim limitation is taught or suggested by Kantor. *Id.* EMC also relies upon Dr. Clark's testimony. *Id.*

PersonalWeb counters that Kantor does not disclose segment identifiers, as recited in claims 10 and 21. PO Resp. 3-14. PersonalWeb also argues that there is insufficient reason to modify Kantor's commands to permit identifying files based on contents signatures. *Id.* at 15-22.

31

Kantor

Kantor describes a method of identifying duplicate files, by using contents signatures that are generated based on the contents of the files, instead of the file names or file locations. Ex. 1004, 2-4, 6-8, 48-49. In particular, Kantor applies a hash function (e.g., a cyclic residue check or cyclic redundancy check ("CRC")) to each file to obtain the contents signature for each file. *Id.* at 6-8, 48-49. For each zip file, Kantor creates zip-file contents signatures by hashing the contents signatures for the files contained within the zip file ("a hash of hashes"). *Id.* at 2, 9. As described by Kantor, this is done by "adding together all the 32_bit CRC's for the files in the zipfile, modulo 2^32, separately adding together their uncompressed file_lengths modulo 2^32, and then arranging the two resulting hexadecimal number as a single structure." *Id*. at 9. Kantor stores the contents signatures and zip-file contents signatures in a master contents-signature list (e.g., CSLIST.SRT). *Id*. at 18.

According to Kantor, contents signatures and zip-file contents signatures are useful for identifying files that have the same contents stored on the electronic bulletin board systems. Ex. 1004, 2 of Preface, 5, 9. For example, when uploading a zip file, the system determines whether that zip file already exists in the system using the zip-file contents signature, and determines whether the inner files of that zip file already exist in the system using the contents signatures for the inner files. *Id*. at 9. Kantor specifically acknowledges the benefits of using contents signatures and zip-file contents signatures to: (1) find files or zip files on the system and delete duplicate

files or zip files uploaded under different names; and (2) determine whether a collection of files that corresponds to one zip file is contained in a larger zip file or spread among several different zip files. *Id.*

Segment Identifiers

Claim 10 recites "said data item comprising a plurality of segments . . . the segment identifier for each particular segment being based, at least in part, on a first given function of the data comprising said particular segment and only the data in said particular segment." Ex. 1001, 41:59-67. Claim 21 also recites a similar limitation. In its petition, EMC asserts that this limitation is met by Kantor's technique of calculating the contents signatures for the inner files of a zip file by applying a function to the inner files. Pet. 43-45 (citing Ex. 1004, Preface, 7-9; Ex. 1009 ¶¶ 34-35).

PersonalWeb responds that Kantor teaches away from the claimed subject matter, because Kantor applies the CRC hash function to *uncompressed* files before they are compressed and packaged into the zip file. PO Resp. 4-13 (citing, *e.g.*, Ex. 2016, 65, 67; Ex. 2020 ¶¶ 25-26). According to PersonalWeb, the CRC hash function is applied to different bit sequences (uncompressed files) than the bit sequences (compressed files) that make up the inner files of the zip file. *Id.*

We are not persuaded by PersonalWeb's arguments, as they are based incorrectly on the assumption that Kantor's inner files of a zip file must be *compressed files*. Rather, we agree with EMC that zip files can have *uncompressed* inner files. Reply 1 (citing Ex. 1092 ¶¶ 6-9; Ex. 1088, 263-64). Indeed, PersonalWeb does not disagree that zip files *are not*

always compressed. Ex. 1088, 263-64. As PersonalWeb's evidence shows, the standard zip-file format, used by Kantor at the time of the invention, defines seven compression methods, which include "Compression method 0" that *does not compress the inner files* when packaging them into a zip file. Ex. 2004, 3; Ex. 1004, 2 of Preface.

Dr. Dewar's reliance on Kantor's statements regarding file compression ratio to support his testimony—"Kantor confirms that the 'files' in the ZIP files described in Kantor are compressed"—is misplaced. Ex. 2020 ¶ 28 (citing Ex. 1004, 2 of Preface, 9, 55). The mere fact that Kantor refers to a compression ratio does not support PersonalWeb's position that the inner files of a zip file must be compressed, because in the situation where "Compression method 0" is used—which *does not compress the inner files*—the file compression ratio is one. Contrary to Dr. Dewar's testimony, those portions of Kantor cited by Dr. Dewar do not require each inner file of a zip file to be compressed. Instead, the cited portions of Kantor merely state that the zip-file contents signature *depends on the contents of the inner files*, and provide examples of items that the zip-file contents signature do not depend upon. *See, e.g.*, Ex. 1004, 2 of Preface ("FWKCS has the special ability to make a 'zipfile contents signature', ('zcs') which is *independent of* . . . the names and dates of files in the zipfile, zipped path information, and file compression ratio."); *id*. at 9 ("This has the desirable property that the resulting zcs *does not depend* on the names of the files, . . . nor on the method nor amount of compression . . . ." (emphasis added)).

As EMC notes, even if Kantor only used compressed inner files, Kantor still would describe the disputed claim limitation—"a first given function of the *data* comprising said particular segment and only the *data* in said particular segment"—as the first given function would include a function that *hashes and compresses the data* in the file. Reply 1 (citing Ex. 1092 ¶ 12). Indeed, nothing in the claim language limits a given function to just *hashing* the data. PersonalWeb does not explain adequately why a given function cannot comprise both *hashing and compressing the data.* Moreover, *compressing* a file merely changes *the format* of the file, but it does not change *the contents* of the file. In other words, both compressed and uncompressed versions of an inner file have the *same contents* (i.e., the data). As discussed above, Kantor's contents signatures are generated based on *the contents* of the files (Ex. 1004, 6-8), and Kantor's zip-file contents signatures depend on *the contents* of the inner files and do not depend on the *format* of the inner files (Ex. 1004, 2 of Preface, 9, 55). Claims 10 and 21 do not place any limitation on *the format* of the plurality of segments.

PersonalWeb's "teaching away" argument is misplaced. Even if Kantor expresses a general preference for applying a function to uncompressed files to obtain contents signatures before they are compressed, that itself does not operate to criticize, discredit, or otherwise discourage investigation into the aforementioned claim limitation. *See DePuy Spine*, 567 F.3d at 1327. As discussed above, Kantor's contents signatures meet

the claim limitation regardless of whether the inner files are compressed or non-compressed.

PersonalWeb also argues that, because a zip file includes additional information (e.g., headers) and Kantor does not obtain a content signature for *headers*, Kantor fails to describe a segment identifier for each segment of the data item. PO Resp. 5-6, 13. That argument is unpersuasive, as it is not commensurate with the scope of claims. *See In re Self*, 671 F.2d 1344, 1348 (CCPA 1982) (It is well established that limitations not appearing in the claims cannot be relied upon for patentability.). Indeed, because claims 10 and 21 each recite the open-ended phrase "comprising" when describing what a data item includes (a "data item *comprising* a plurality of segments"), a data item is not limited to just a plurality of segments, and may include additional information (e.g., headers). Furthermore, as discussed above, the claim term a "data item" includes "a *portion* of a file." Consequently, the inner files meet the claim term "a plurality of segments," and Kantor's technique of generating contents signatures by applying a function to the inner files describes the disputed claim limitation.

Contrary to PersonalWeb's argument that Kantor merely "reads" the CRC values (PO Resp. 9-10), Kantor determines a zip-file contents signature by "adding together all the 32_bit CRC's for the files in the zip file, modulo 2^{32}, separately adding together their uncompressed_file_lengths modulo 2^{32}, and then arranging the two resulting hexadecimal numbers as a single structure." Ex. 1004, 9. Dr. Clark testifies that addition module 2^{32} is a well-known simple hashing function that uses addition to calculate a value

for a file based on the contents of the file. Ex. 1009 ¶ 35 (citing Ex. 1011).
Dr. Clark's testimony is consistent with Kantor's disclosure that the
resulting zip-file contents signature depends on the contents of the inner files
and "does not depend on the names of the files, the dates of the files, the
order in which they appear in the zip file, nor on the method nor amount of
compression, nor does it depend on comments." Ex. 1004, 3, 9.

Given Kantor's express disclosure and the evidence before us, we
determine that EMC has demonstrated sufficiently that Kantor describes
segment identifiers for a plurality of segments as recited in claims 10 and 21.

Identifying files using contents signatures

Claim 10 recites "using at least one of said segment identifiers . . .
requesting at least one particular segment of said plurality of segments . . .
obtaining said particular segment from said at least one of a plurality of
computers in said network of computers." Ex. 1001, 42:6-12.

In its petition, EMC recognizes that the users typically request files
based on the file names. Pet. 45. Nonetheless, EMC asserts that a person
having ordinary skill in the art would have found it obvious to modify the
electronic Bulletin Board Systems commands, including the download and
read commands, to identify files using contents signatures or zip-file
contents signatures, instead of file names. *Id*. at 46 (citing Ex. 1009 ¶ 41).
According to EMC, "this would facilitate integrity checking by more
precisely specifying the file of interest by its content, and thus improve
accuracy." *Id*. Dr. Clark testifies that such a modification would provide a

more efficient and context-free means for accessing and sharing files.
Ex. 1009 ¶ 41.

PersonalWeb counters that it would not have been obvious to modify
Kantor so that the read and download requests would accept contents
signatures to identify files. PO Resp. 15-22. In particular, PersonalWeb
argues that Kantor teaches away from replacing conventional file names
with contents signatures for identifying files, because "Kantor intentionally
designed his contents-signatures so that certain different files would have the
same signature." PO Resp. 16-18 (citing Ex. 1004, 3, 51; Ex. 2020 ¶¶ 44-
46). PersonalWeb also alleges that Kantor fails to teach or suggest the
alleged modification, and fails to provide any suggestion or motivation for
the alleged modification. *Id*. at 18-22 (citing Ex. 2020 ¶¶ 47-48).
PersonalWeb submits that Kantor does not disclose any problems with the
use of conventional file names for the read and download requests. *Id*. at 22.

We are not persuaded by PersonalWeb's arguments. First,
PersonalWeb's teaching away argument is misplaced, as it fails to recognize
that the cited portion of Kantor specifically explains that the different files
that allegedly have the same signature files also have the *same contents*. *See*
Ex. 1004, 3 ("[T]he same file contents . . . will have the same zipfile
contents signature."). In fact, that is one of the reasons why using contents
signatures or zip-file contents signature, instead of file names, to identify
files is more accurate. Ex. 1004, Preface, 5, 9. Notably, files that have the
same contents would be identified as duplicates, and files that have *different*
contents would be identified as different files, regardless of whether they

have different file names. *Id.* As Kantor notes, finding and deleting duplicate files would improve system efficiency. *Id.*

Further, PersonalWeb's argument that Kantor does not teach or suggest the alleged modification is unpersuasive, because an obviousness analysis "need not seek out precise teachings directed to the specific subject matter of the challenged claim, for a court can take account of the inferences and creative steps that a person of ordinary skill in the art would employ." *KSR*, 550 U.S. at 418. PersonalWeb's argument overlooks "the fundamental proposition that obvious variants of prior art references are themselves part of the public domain." *Translogic,* 504 F.3d at 1259. Moreover, we observe that the asserted ground of unpatentability is based on the *combination* of Kantor's teaching of using contents signatures to identify files with Kantor's teaching of requesting files. It is well settled that nonobviousness cannot be established by attacking each prior art teaching individually where, as here, the ground of unpatentability is based upon a combination of different teachings in the prior art. *See In re Keller*, 642 F.2d 413, 426 (CCPA 1981). Rather, the test for obviousness is whether the combination of prior art teachings, taken as a whole, would have suggested the patentees' invention to a person having ordinary skill in the art. *See In re Merck & Co.*, 800 F.2d 1091, 1097 (Fed. Cir. 1986).

As to PersonalWeb's arguments that Kantor does not provide a motivation for the modification (PO Resp. 18; Ex. 2020 ¶ 47), a rationale to combine the prior art teachings does not have to be found explicitly in the prior art, itself. *See In re Kahn*, 441 F.3d 977, 987 (Fed. Cir. 2006) (A

"motivation to combine the relevant prior art teachings does not have to be found explicitly in the prior art."). We also are not persuaded by PersonalWeb's argument that there would have not been a logical reason to modify Kantor in the manner alleged by EMC, other than impermissible hindsight (PO Resp. 22). As discussed above, EMC asserts that it would have been obvious to modify the read and download commands to identify files using contents signatures instead of file names. Pet. 46 (citing Ex. 1009 ¶ 41). EMC takes the position that "this would facilitate integrity checking by more precisely specifying the file of interest by its content, and thus improve accuracy." *Id.* Dr. Clark testifies that such a modification would provide a more efficient and context-free means for accessing and sharing files. Ex. 1009 ¶ 41. EMC's position and Dr. Clark's testimony are consistent with Kantor's disclosure that using contents signatures, instead of file names, to find and delete duplicate files would increase system efficiency by reducing storage cost and system time for locating and managing files. Ex. 1004, Preface, 5, 9, 205-206. As such, we conclude that EMC has articulated a sufficient reason to combine the teachings of Kantor.

We are not persuaded by PersonalWeb's arguments that the proposed modification is not enabled and that EMC fails to explain how the proposed modification could have been carried out to yield a predictable result. PO Resp. 19-22. EMC specifically explains that Kantor's Precheck and Lookup operations provide examples of user commands that utilize contents signatures. Pet. 46 (citing Ex. 1004, 97, 173; Ex. 1009 ¶¶ 37, 39, 41). For instance, Kantor describes the Precheck operation as a software utility

running on the electronic Bulletin Board Systems for identifying files that already uploaded in the system by using their contents signatures. Pet. 46-47 (citing Ex. 1004, 173). Dr. Clark explains that Kantor's Lookup operation permits users to submit a request containing a contents signature to determine where the corresponding file is located on the system. Ex. 1009 ¶ 41 (citing Ex. 1004, 96-97). As Dr. Clark also notes, Kantor's "i" function provides users with the capability to submit contents-signature search requests to find files on the system that contain material related to a user's file, by obtaining the contents signatures for the inner files of the zip files that contain the related material. *Id.* (citing Ex. 1004, 96-97). Dr. Clark further testifies the system as modified would have utilized one of those contents signatures for the inner files in a download request to obtain the particular inner file that is associated with the contents signature. *Id.* Upon review of the parties' contentions and supporting evidence, we credit Dr. Clark's testimony as it is consistent with Kantor's disclosure. We also agree with EMC that Dr. Clark merely relies on the disclosure of Kantor (Ex. 1004, 96-97), and not LOOKUP.DOC and PRECHECK.DOC files as alleged by PersonalWeb. For the foregoing reasons, we determine that EMC has explained sufficiently how the proposed modification could have been carried out to yield a predictable result.

In addition, PersonalWeb agrees that, at the time of the invention, users on the electronic Bulletin Board Systems had the capability to request a file using the *file name*. PO Resp. 18. In light of Kantor, a person of ordinary skill in the art would have recognized how to calculate contents

signatures and zip-file contents signatures and how to use them to identify files. *See, e.g.*, Ex. 1004, Preface, 5-9. A person with ordinary skill in the art also would have appreciated the benefit of using contents signature and zip-file contents signatures that are generated based on the contents of the files, rather than *file names*, for identifying files accurately. *Id.* The mere substitution of contents signatures and zip-file contents signatures for *file names* in read and download requests predictably uses prior art elements according to their established functions. Such a substitution is an obvious improvement. *See KSR*, 550 U.S. at 417 (The simple substitution of one known element for another is likely to be obvious if it does no more than yield predictable results.). Moreover, PersonalWeb has not provided sufficient evidence that such a substitution is beyond the level of a person with ordinary skill in the art. *See Leapfrog Enters., Inc. v. Fisher-Price, Inc.*, 485 F.3d 1157, 1162 (Fed. Cir. 2007).

Conclusion

We also are not persuaded by PersonalWeb's evidence of non-obviousness, because it fails to establish the required nexus, as discussed above. For the foregoing reasons, we determine that EMC has demonstrated by a preponderance of the evidence that claims 10 and 21 are unpatentable over Kantor.

G. Claim 34 – Obvious Over Kantor and Langer

EMC asserts that claim 34 is unpatentable under 35 U.S.C. § 103(a) as obvious over Kantor and Langer. Pet. 42-50 (citing Ex. 1003, 3-5; Ex. 1004,

Preface, 7-11, 48, 51-55, 96-97, 173-74). EMC maintains that Kantor, as understood by a person with ordinary skill in the art, renders claim 34 obvious. *Id*. at 42-49. However, EMC relies upon Langer to teach a cryptographic hash function in case the claim term "a True Name of the data" is limited to use of such a function. *Id*. at 49-50. EMC also relies upon its explanations regarding the obviousness ground of unpatentability based on Kantor, itself, as to claims 10 and 21, and Dr. Clark's testimony. *Id*. at 42-49 (citing Ex. 1004, Preface, 7-11, 48, 51-55, 96-97, 173-74).

PersonalWeb counters that Langer does not cure the deficiencies of Kantor. PO Resp. 22-39. PersonalWeb further argues that the combination of Kantor and Langer does not disclose all of the limitations recited in claim 34. *Id*. PersonalWeb also relies upon some of the same arguments presented with respect to claims 10 and 21. *See, e.g., id.* at 35-37.

Based on the evidence before us, we are not persuaded by PersonalWeb's arguments. In the analysis below, we focus on the deficiencies alleged by PersonalWeb as to the ground of unpatentability based on the combination of Kantor and Langer, and we address each of PersonalWeb's argument in turn.

Dividing a data item into a plurality of segments

Claim 34 recites "dividing a particular data item into a plurality of segments; . . . determining a plurality of segment identifiers by . . . determining a corresponding segment identifier for each particular segment of said plurality of segments." Ex. 1001, 45:5-9. PersonalWeb argues that Kantor fails to disclose that limitation. PO Resp. 23-44.

First, PersonalWeb argues that Kantor does not determine segment
identifiers for *all segments* of a zip file. *Id*. at 23-34 (citing Ex. 2020
¶¶ 50-52). We are not persuaded. As discussed above, an example of a data
item is a *portion* of a file. As Dr. Clark testifies (Ex. 1092 ¶¶ 28-33),
"a sequence of bits" does not require that the bits must be *contiguous*. We
also observe that a large database file—a data item—may be stored across
different storage systems or memories. *See, e.g.*, Ex. 1005, 14:53-15:8;
Ex. 1003, 5. Therefore, the "dividing" limitation does not require
necessarily the zip file, *in its entirety*, to be divided into a plurality of
segments. In other words, a plurality segments may include merely the inner
files of a zip file, excluding headers and other information about the data.

Next, PersonalWeb alleges that Kantor intentionally designed his
contents signatures so that different files would have the same signature. PO
Resp. 24. As discussed above, that argument is unpersuasive because it fails
to recognize that the cited portion of Kantor specifically explains that the
different files that allegedly have the same signature also have the *same
contents*. *See* Ex. 1004, 3 ("[T]he same file contents . . . will have the same
zipfile contents signature.").

According to PersonalWeb, the disputed claim limitation expressly
requires that the step of *determining* the segment identifiers take place *after*
the "*dividing*" step. PO Resp. 35 (citing Ex. 2020 ¶¶ 26-32, 62-63). Based
on that interpretation, PersonalWeb alleges that Kantor does not meet that
limitation, because Kantor determines the *CRC values* before the files are
packaged into a zip file. *Id*.

PersonalWeb's argument is not persuasive, as it does not consider the situation in which the new file to be uploaded to the system is already a zip file (Ex. 1004, 9). In that situation, the inner files of such a zip file are packaged in a zip file *before* the CRC values are calculated.

Even applying PersonalWeb's interpretation of the disputed claim limitation and its reading of Kantor, we determine that EMC has established sufficiently that Kantor's *contents signatures* render the disputed claim limitation obvious. We observe that a *CRC value* is not the same as the *contents signature* (segment identifier) of an inner file. That is an important distinction. The *contents signature* (segment identifier) for each inner file is generated by using the *CRC value and the length value* of the file. Ex. 1004, 8. Kantor discloses that the system looks inside (i.e., *unzips*) the zip file, and then uses the information (e.g., the *CRC values*) in the zip file to generate the *contents signatures* (segment identifiers) for the inner files in the zip file. Ex. 1004, 48. According to Kantor, this is a relatively quick operation, as the zip file already contains the *CRC values*. *Id.* Therefore, even if, as alleged by PersonalWeb, *the CRC values* are calculated before the files are packaged into a zip file, Kantor explicitly discloses that the *contents signatures* (segment identifiers) for the inner files are calculated *after* the inner files are packaged into a zip file. Ex. 1004, 8, 48-49.

For the foregoing reasons, we determine that EMC has demonstrated sufficiently that the combination of Kantor and Langer teaches or suggests the "dividing" claim limitation.

Segment identifier being a True Name

Claim 34 recites "the segment identifier for each particular segment being a True Name of the data comprising said particular segment." Ex. 1001, 45:10-12. As discussed above, the claim term "True Name" is construed as "a substantially unique alphanumeric label for a particular data item." PersonalWeb alleges that the combination of Kantor and Langer does not describe that limitation. PO Resp. 35-37. PersonalWeb relies upon the same arguments presented with respect to claims 10 and 21. *Id.* As discussed above, those arguments are unpersuasive.

Determining a True Name of the second data item

Claim 34 recites "forming a second data item comprising said plurality of segment identifiers, . . . in response to a request to access said data item, said request comprising said data item identifier, providing at least said second data item." Ex. 1001, 46:1-10.

In its patent owner response, PersonalWeb alleges that the combination of Kantor and Langer does not describe the aforementioned limitation. PO Resp. 37-39. Specifically, PersonalWeb assumes that the second data item is made up of Kantor's *CRC values*, and based on that assumption, it argues the following: (1) Kantor fails to disclose arranging the *CRC values* in a sequence of bits (*id*. at 38); (2) the zip-file contents signature is not a True Name of a sequence of the *CRC values* (*id*. at 39); and (3) Kantor fails to disclose providing all the *CRC values* in response to a

request to access the data item that includes the data item identifier (*id*.).
We are not persuaded by PersonalWeb's arguments.

We instead agree with EMC that PersonalWeb's assumption that the
second data item is made up of Kantor's *CRC values* is incorrect. Reply 7.
PersonalWeb's argument, once again, ignores the important distinction
between the *CRC values* and the *contents signatures* for the inner files. As
EMC notes, Kantor explicitly discloses that the *contents signatures* of the
inner files are stored in a master contents-signature list. Pet. 45 (citing
Ex. 1004, 18). Further, Dr. Clark explains that Kantor's Lookup *command*
(Ex. 1004, 173) which uses the "remote Inquiries" option (the "i" function)
provides users with the capability to submit contents-signature search
requests to find files on the system that contain material related to a user's
file. Ex. 1009 ¶ 41 (citing Ex. 1004, 96-97); *see also* Pet. 47. Dr. Clark also
testifies that, in response to a Lookup request including a zip-file contents
signature when using the "y form of the TEST" function, the system
provides the user *the full set of contents signatures* (a second data item) for
all the inner files in each of the zip files in which the specific file appears.
Ex. 1092 ¶ 41 (citing Ex. 1004, 96-98). We credit Dr. Clark's testimony as
it is consistent with the explicit disclosure of Kantor.

Conclusion

We also are not persuaded by PersonalWeb's evidence of non-
obviousness, because it fails to establish the required nexus, as discussed
above. For the foregoing reasons, we determine that EMC has demonstrated

by a preponderance of the evidence that claim 34 is unpatentable over the combination of Kantor and Langer.

*H. Claims 10 and 21 – Obvious Over Woodhill and Fischer*

EMC asserts that claims 10 and 21 are unpatentable under 35 U.S.C. § 103(a) as obvious over Woodhill and Fischer. Pet. 50-57 (citing Ex. 1009 ¶¶ 44-59). EMC acknowledges that Woodhill's disclosure of restoring a file does not use a hash of the granule identifiers to identify a database file. Pet. 50-55. Nevertheless, EMC indicates that using a "*hash of hashes*" technique for identifying database or compound files was known in the art at the time of the invention, as evidenced by Fischer (Ex. 1036, 7:49-8:38). Pet. 56.

PersonalWeb counters that the combination of Woodhill and Fischer does not describe certain claim limitations, and it would not have been obvious to combine Woodhill and Fischer in the manner asserted by EMC. PO Resp. 47-54. Upon review of the evidence on record, we are not persuaded by PersonalWeb's arguments.

Woodhill

Woodhill discloses a system for distributed storage management using binary object identifiers. Ex. 1005, 1:11-17. The system includes a remote backup file server in communication with a plurality of local area networks. *Id*. Woodhill's system includes a Distributed Storage Manager ("DSM") program for building and maintaining a file database. *Id*. at 3:44-49, fig. 3. The DSM program views a file as a collection of data streams, and divides each data stream into one or more binary objects. *Id*. at 4:13-23, 7:40-43,

fig. 5A, item 132. For each binary object being backed up, a binary object identification record is created in a file database and includes a binary object identifier to identify each binary object uniquely. *Id.* at 7:60-8:1, 8:33-34. Binary object identifiers are calculated based on the contents of the data instead of from an external and arbitrary source, such that the binary object identifier changes when the contents of the binary object changes. *Id.* at 8:57-62, 8:40-42. The DSM program also utilizes a technique of subdividing the large database files into granules and then tracks changes from the previous copy of the granules. *Id.* at 14:53-65. This technique is used to reduce the amount of data that must be transmitted to the remote backup file server. *Id.* at 15:4-8. Figure 5I of Woodhill, reproduced below, illustrates the process of restoring a file to a previous version:

FIG. 5I
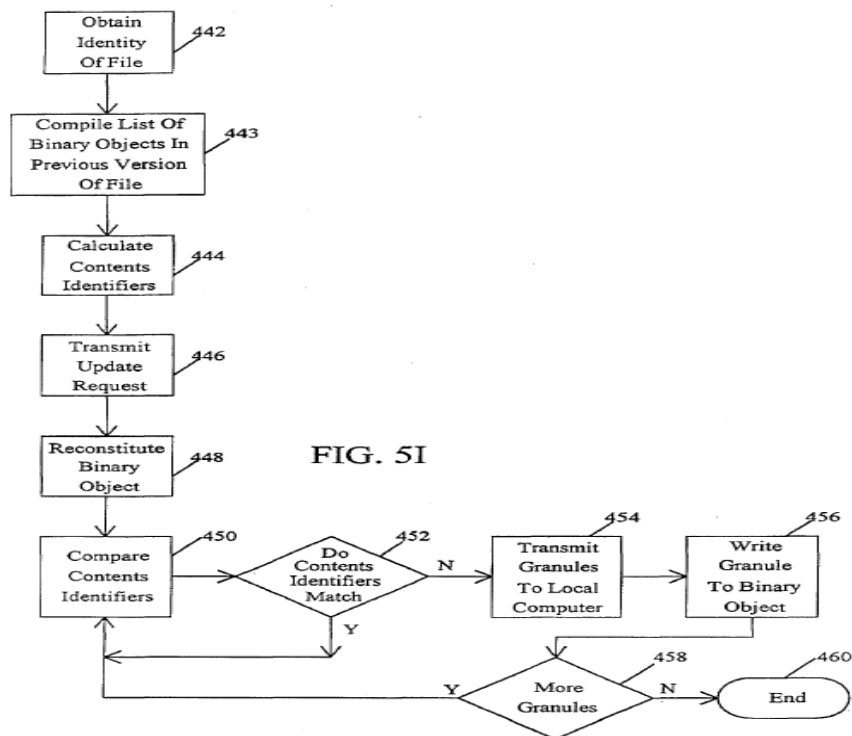
As shown in Figure 5I of Woodhill, in response to a user's request to restore a file, the DSM program restores the previous version of the binary object by retrieving the granules from the remote server. *Id*. at 17:18-18:9. The DSM program uses the previous version of granule contents identifiers to determine the location of the granules. *Id*. at 17:50-55, fig. 5I, box 450. It compares the previous version of contents identifiers with the contents identifiers for the granules within the current version of the file. *Id*. The DSM program locates the granule on the local computer when the contents identifiers match. *Id*. at 17:58-60. If the contents identifiers do not match, the DSM program locates the granule on the remote server and transmits the granule from the remote server to the local computer. *Id*. at 17:60-64, fig. 5I, box 454. After the granules that are located on the remote server have been transferred to the local computer, the file on the local computer is restored to its previous version. *Id*. at 18:6-9.

Using a segment identifier to request a segment

Claim 10 recites "using at least one of said segment identifiers . . . requesting at least one particular segment of said plurality of segments that comprise said data item." Ex. 1001, 42:6-9. Claim 21 recites a similar limitation.

In its petition, EMC relies upon Woodhill's disclosure of restoring a file to meet this limitation. Pet. 54-55 (citing Ex. 1005, 17:18-18:9). Specifically, EMC asserts that Woodhill discloses an *update request* to restore a current version of a binary object to a prior version of a binary object, which includes the binary object identifier for the prior version of the

binary object. *Id.* at 54 (citing Ex. 1005, 7:60-8:4, 17:17-50; Ex. 1009 ¶¶ 48, 50-52). According to EMC, in response to the *update request*, the DSM program uses the contents identifiers (segment identifiers) for the granules to obtain a granule (segment) from the remote backup file server for the local computer. *Id.* at 55 (citing Ex. 1005, 17:50-18:9; Ex. 1009 ¶¶ 46, 49, 53).

PersonalWeb counters that Woodhill does not disclose the "request" limitation. PO Resp. 47-51 (citing Ex. 2020 ¶¶ 88-93). In particular, PersonalWeb asserts that Woodhill does not use contents identifiers to request granules of a binary object, because "a contents identifier is never provided in any 'request' for a particular granule." *Id.* at 48-50 (citing Ex. 2020 ¶¶ 90-93). According to PersonalWeb, the contents identifiers are compared to determine whether to transmit granules, but are not used for *requesting* granules. *Id.* at 51 (citing Ex. 1005, 17:51-65; Ex. 2020 ¶ 92).

We are not persuaded by PersonalWeb's arguments, as they are not commensurate with the scope of claims 10 and 21. Rather, we agree with EMC that the claim limitation does not require a content identifier (segment identifier) to be *provided in a request* for a particular granule (segment). *See* Reply 11-12 (citing Ex. 1005, 17:18-46; Ex. 1009 ¶¶ 48-54). Claims 10 and 21 merely require *using* a segment identifier to request a particular segment. In fact, PersonalWeb's expert acknowledges that "the contents identifiers are *used* to determine which granules have changed via the comparison," and to identify which granule is to be transmitted from the remote backup file server to the local computer. PO Resp. 49-50; Ex. 2020 ¶ 92 (emphasis added). Further, Woodhill discloses that an *update request* includes the

51

binary object identifiers for the binary object of the previous version of the file, as well as granule contents identifiers (segment identifiers) for each granule of the current version (segment). Ex. 1005, 17:36-46; Ex. 1088, 185. As discussed above, during the comparison of the granule contents identifiers, if the contents identifiers do not match, the DSM program locates the particular granule on the remote server and transmits the granule from the remote server to the local computer. *Id*. at 17:60-64, fig. 5I, box 454. In other words, the DSM program requests the particular granule using its contents identifiers from the remote server.

We also are not persuaded by PersonalWeb's argument that "there is no disclosure in Woodhill of using a Binary Object Identifier 74 to obtain a plurality of contents identifiers." PO Resp. 51 (citing Ex. 2020 ¶ 93). Instead, we agree with EMC that the binary object identifier is used to obtain its corresponding granule content identifiers. Reply 12 (citing Ex. 1005, fig. 3; Ex. 1009 ¶ 52; Ex. 1088, 185, 196-197). Woodhill discloses that, in response to the user's request to restore a file to the previous version, the DSM program compiles a list of all binary objects comprising the current version of the user-specified file from the file database, which includes the binary object identifiers of all the binary objects for the file. Ex. 1005, 17:27-36, fig. 3. The DSM program then calculates the *contents identifiers* for granules within the current version of each binary object. *Id*. at 17:36-40. Given the evidence before us, we determine that EMC has demonstrated sufficiently that the combination of Woodhill and Fischer would have rendered the "request" claim limitation obvious.

52

First identifier is based on a second function of segment identifiers

Claim 10 recites "said first identifier is based, at least in part, on a second given function of the plurality of segment identifiers." Ex. 1001, 42:2-5. Claim 21 recites a similar limitation. EMC acknowledges that Woodhill's disclosure of restoring a file does not use a hash of the granule identifiers to identify the database file that contains the granules. Pet. 50-55. Nevertheless, EMC indicates that using a "*hash of hashes*" technique for identifying database or compound files was well known, as evidenced by Fischer (Ex. 1036, 7:49-8:38). *Id.* at 56. EMC contends that a person of ordinary skill in the art would have modified Woodhill's file restoring process by calculating the identifier for the large database file based on a function of the granule contents identifiers ("a hash of hashes"), as taught by Fischer, because this would improve the efficiency and performance of Woodhill's data processing for restoring a file. *Id.* at 57 (citing Ex. 1009 ¶ 59). As support, Dr. Clark testifies that "if only a few granules are changed, it is faster to compute a hash of the granule identifiers (rather than of the entire binary object) because the previously calculated granule identifiers could be re-used." Ex. 1009 ¶ 59. In response, PersonalWeb advances three arguments to support its contention that it would not have been obvious to modify Woodhill. PO Resp. 51-54.

1. *Hashing contents identifiers*

PersonalWeb argues that the binary object identifiers could not have been generated based on contents identifiers, because the contents identifiers

do not exist when the binary object identifiers are calculated. *Id*. at 52 (citing Ex. 2020 ¶ 96). However, EMC counters that PersonalWeb's assumption that the binary object identifiers must be calculated *before* contents identifiers are determined is incorrect, because Woodhill does not impose such a requirement. Reply 12-13 (citing Ex. 1092 ¶¶ 68-69).

We agree with EMC that Woodhill does not require any specific order for calculating binary object identifiers and contents identifiers. The portion of Woodhill cited by PersonalWeb does not support its assumption. PO Resp. 52 (citing Ex. 2020 ¶ 96; Ex. 1005, 17:44). As Dr. Clark testifies, the calculation of contents identifiers does not depend on a binary object identifier. Ex. 1092 ¶ 68 (Ex. 1005, 5:12-9:28, 14:52-18:9). Dr. Clark also testifies that the cited portion of Woodhill merely demonstrates that the binary object identification records for the *previous version* of the binary object exist at the time an update request for restoring the previous version is made, and the contents identifiers calculated in step 444 of Woodhill's Figure 5I are for the granules within the *current version* of each binary object as it exists on the local computer. Ex. 1092 ¶ 69. We credit Dr. Clark's testimony as it is consistent with Woodhill's disclosure. *See* Ex. 1005, 5:12-9:28, 14:52-18:9, 17:18-50.

2. *Teaching away argument*

PersonalWeb argues that Woodhill teaches away from the modification. PO Resp. 52-53 (citing Ex. 2020 ¶ 97). According to PersonalWeb, Woodhill teaches that Fischer's technique is undesirable, because Woodhill emphasizes that the binary object identifier is calculated

from the *contents* of the data instead of from an external and arbitrary source, whereas Fischer calculates the "fileHash" *using external and arbitrary sources*. *Id*. (citing Ex. 1005, 8:40-42; Ex. 1036, 8:4-55). EMC counters that PersonalWeb's argument is based on the incorrect assumption that the record identifiers "K" of Fischer's fileHash must be related to *external* information. Reply 13. Dr. Clark testifies that PersonalWeb's reliance on the statement of Fischer (Ex. 1036, 5:55-58) is incorrect, because an "employee number" corresponds to exactly one employee record and, therefore, is *neither external nor arbitrary*. Ex. 1092 ¶ 71.

We agree with EMC that Fischer does not require the record identifiers to be based on *external* information. In fact, in the same sentence relied upon by PersonalWeb, Fischer discloses that "a record number [] K_i *may* be any indexing value." Ex. 1036, 5:55-58 (emphasis added). Even if Fischer's technique requires a calculation using an *external* source, obviousness does not require that all of the features of the secondary reference be bodily incorporated into the primary reference. *In re Etter*, 756 F.2d 852, 859 (Fed. Cir. 1985) (en banc); *Keller*, 642 F.2d at 425.

In any event, EMC's proposed modification does not change Woodhill's process for calculating the contents identifiers which are based on the *contents* of the granules. Pet. 55-57. By applying "a hash of hashes" technique, the binary object identifiers for the database file would be calculated using a hash function (second function) against the *contents identifiers* of the granules associated with the binary objects. *Id*. Such

binary object identifiers still would be based on the *contents* of the binary object, as Woodhill's "key notion" statement suggests. Ex. 1005, 8:38-42.

We are not persuaded by PersonalWeb's argument that Woodhill is concerned with uniquely identifying a binary object, whereas Fischer is concerned with security. The mere fact that the two references have different objectives does not mean that a person with ordinary skill in the art would not combine their teachings. *Heck*, 699 F.2d at 1333 ("The use of patents as references is not limited to what the patentees describe as their own inventions or to the problems with which they are concerned." (citation and internal quotation marks omitted)); *EWP Corp.*, 755 F.2d at 907.

3. *Argument regarding inoperability*

PersonalWeb asserts that the alleged combination would have rendered Woodhill's system inoperable for its intended purpose. PO Resp. 53-54 (Ex. 2020 ¶ 98). Specifically, PersonalWeb argues that many parts of Woodhill's system rely on binary object identifiers to detect changes in binary objects that are not granularized, and, thus, one of ordinary skill in the art would not have modified Woodhill's binary object identifiers to base them on granule contents identifiers because this would have resulted in Woodhill's system being inoperative. *Id.*

However, EMC's proposed modification is limited to Woodhill's restoring process in which a binary object has been divided into a plurality of *granules*, and is not limited to the non-granularization situations, as alleged by PersonalWeb. Pet. 54-56 (citing Ex. 1005, 17:7-50, 17:60-18:4). The binary object identifiers for the database file being restored would be

calculated based on a function of *the content identifiers of the granules.*
Pet. 56. Moreover, as Dr. Clark explains, Woodhill recognizes that a binary
object identifier may be calculated in various ways. Ex. 1009 ¶ 72 (citing
Ex. 1005, 8:38-40). Dr. Clark further testifies that, regardless of whether a
binary object identifier is calculated directly from the contents of the binary
object, *or* it is calculated as a function of granule contents identifiers for the
granules associated with the binary object, the binary object identifier is
based on *the contents of the binary object*. *Id.* Therefore, we are not
persuaded that EMC's proposed modification would render Woodhill's
system inoperable for its intended purpose. *Keller*, 642 F.2d at 425 ("The
test for obviousness is not whether the features of a secondary reference may
be bodily incorporated into the structure of the primary reference . . . .").

Conclusion

We also are not persuaded by PersonalWeb's evidence of non-
obviousness, because it fails to establish the required nexus, as discussed
above. For the foregoing reasons, we determine that EMC has demonstrated
by a preponderance of the evidence that claims 10 and 21 are unpatentable
over Woodhill and Fischer.

I. EMC's Motion to Exclude

EMC seeks to exclude: (1) three license agreements (Exs. 2010-12);
(2) Mr. Kevin Bermeister's declarations (Exs. 2009, 2022) relating to those
license agreements; and (3) Mr. Todd Thompson's declaration (Ex. 2014).
Paper 59 ("Pet. Mot."). PersonalWeb filed the license agreements and

Mr. Bermeister's declarations as evidence of non-obviousness to rebut
EMC's assertion that claims 10, 21, and 34 would have been obvious over
the various combinations of Langer, Kantor, Woodhill, and Fischer. PO
Resp. 54. As to Mr. Thompson's declaration, PersonalWeb proffered that
evidence to support its assertion that Kantor was not made sufficiently
accessible to an interested person. *Id*. at 58-60. PersonalWeb opposes
EMC's motion to exclude. Paper 63. In response, EMC filed a reply to
PersonalWeb's opposition to its motion to exclude. Paper 66.

With respect to the license agreements and Mr. Bermeister's
declarations, EMC argues that they are irrelevant under Federal Rule of
Evidence 402[4] and highly prejudicial, confusing, and misleading under
Federal Rule of Evidence 403. Pet. Mot. 1-13. As to Mr. Thompson's
declaration, EMC argues that it should be excluded under Federal Rule of
Evidence 402. *Id.* at 14-15. Specifically, EMC alleges the following:
(1) Mr. Thompson does not possess the skill of a person of ordinary skill in
the art (*id.* at 14 (citing Ex. 1086, 13-14)); (2) Mr. Thompson did not use
compatible software from the relevant time period (*id.* (citing Ex. 1086,
40-41; Ex. 2014, 4, 6)); and (3) Mr. Thompson did not follow the
instructions provided with the zip file (*id.* (citing Ex. 1086, 32-35)).

The current situation does not require us to assess the merits of
EMC's motion to exclude. As discussed above, even without excluding
PersonalWeb's supporting evidence, we have determined that Kantor is a

---

[4] As stated in 37 C.F.R. § 42.62, the Federal Rules of Evidence generally
apply to *inter partes* reviews.

"printed publication" under 35 U.S.C. § 102(b) and EMC has demonstrated by a preponderance of the evidence that claims 10, 21, and 34 are unpatentable over the various combinations of Langer, Kantor, Woodhill, and Fischer.

Accordingly, EMC's motion to exclude evidence is *dismissed* as moot.

J. PersonalWeb's Motion to Exclude

PersonalWeb seeks to exclude the following items of evidence: (1) Kantor (Ex. 1004) and Langer (Ex. 1003); (2) certain documents that corroborate the knowledge and recollections of EMC's witnesses (Exs. 1050-1052, 1055-1058, 1061-1064, 1073, 1077, 1078, 1083-1085), and the portions of testimony regarding these documents; (3) the declarations of Messrs. Sussell, Sadofsky, and Moore (Exs. 1053, 1081, 1091, 1059), and Mr. Sadofsky's deposition (Ex. 2013, 30, 66); and (4) Clark's rebuttal declaration (Ex. 1092). Paper 55 ("PO Mot.").

EMC opposes PersonalWeb's motion to exclude. Paper 64 ("Opp.").
In response, PersonalWeb filed a reply to EMC's opposition to its motion to exclude. Paper 67 ("PO Reply"). For the reasons stated below, PersonalWeb's motion to exclude is *denied*.

1. Kantor and Langer

PersonalWeb alleges that Kantor and Langer should be excluded as unauthenticated and inadmissible hearsay under Federal Rules of Evidence 901 and 902. PO Mot. 1-6. In particular, PersonalWeb argues that "[n]o witness of record has personal knowledge of Kantor or Langer existing prior

to [the critical date], and electronic data such as Kantor and Langer is inherently untrustworthy because it can be manipulated from virtually any location at any time." *Id*. at 1. According to PersonalWeb, the dates provided by Kantor and Langer are inadmissible hearsay because Kantor and Langer are not self-authenticating. *Id*.

EMC argues that Kantor and Langer have been authenticated under Federal Rules of Evidence 901, and that the documents are not hearsay, because they are being offered for what they describe—not for the truth of their disclosures. Opp. 1-10. In particular, EMC disagrees with PersonalWeb that the documents cannot be authenticated without direct testimony from a witness with personal knowledge that the documents existed prior to the critical date. *Id*. at 1. EMC asserts that it need "only produce evidence 'sufficient to support a finding' that the reference 'is what the proponent claims it is.'" *Id*. (citing Fed. R. Evid. 901(a)). EMC also contends that testimony from Messrs. Sussell, Sadofsky, and Moore provides sufficient evidence to authenticate Kantor and Langer. *Id*. at 1-10 (citing Exs. 1053, 1081, 1091, 1059).

In its reply, PersonalWeb argues that the Federal Rules of Evidence identified by EMC are not applicable to Kantor and Langer, because Messrs. Sussell, Sadofsky, and Moore did not post or review the documents prior to the critical date. PO Reply 1-5. PersonalWeb also alleges that the authenticity of Kantor and Langer is suspicious, as electronic data is inherently untrustworthy and there is no chain of custody. *Id*.

We are not persuaded that Kantor and Langer should be excluded. At the outset, we disagree with PersonalWeb's position that a witness cannot authenticate a document, unless the witness is the author of the document or the witness has reviewed the document prior to the critical date. Federal Rule of Evidence 901(a) states that the authentication requirement is satisfied if the proponent presents "evidence sufficient to support a finding that the item is what it proponent claims it is." Therefore, neither a declaration from the author, nor evidence of someone actually viewing the document *prior to critical date*, is required to support a finding that the document is what it claims to be. *See Hall*, 781 F.2d at 899 (concluding that "competent evidence of the general library practice may be relied upon to establish an approximate time when a thesis became accessible"); *Wyer*, 655 F.2d at 226 (Notwithstanding that there is no evidence concerning actual viewing or dissemination of any copy of the Australian application, the court held that "the contents of the application were sufficiently accessible to the public and to persons skilled in the pertinent art to qualify as a 'printed publication.'").

Further, it is well settled that an uninterrupted chain of custody is not a prerequisite to admissibility, but rather gaps in the chain go to weight of the evidence. *U.S. v. Wheeler*, 800 F.2d 100, 106 (7th Cir. 1986); *see also U.S. v. Aviles*, 623 F.2d 1192, 1198 (7th Cir. 1980) ("If the trial judge is satisfied that in reasonable probability the evidence has not been altered in any material respect, he may permit its introduction." (citation omitted)). There is a strong public policy for making all information filed in a quasi-

judicial administrative proceeding available to the public, especially in an
inter partes review, which determines the patentability of a claim or claims
in an issued patent. It is within the Board's discretion to assign the
appropriate weight to be accorded to evidence.

Although Messrs. Sussell, Sadofsky, and Moore personally did not
review Kantor or Langer *prior to the critical date*, they nevertheless have
sufficient personal knowledge and working experience to provide competent
testimony to establish the publication and authentication of the documents.
See Hall, 781 F.2d at 899; *Wyer*, 655 F.2d at 226; *Bayer*, 568 F.2d at 1361.
Notably, Mr. Sussell, the co-founder and system operator of the Invention
Factory Bulletin Board System, testifies that Dr. Kantor released his
software on the Invention Factory Bulletin Board System in the 1980s, and
his system continuously utilized and hosted current versions of the software
and user manuals. Ex. 1053 ¶¶ 3, 13, 15. Mr. Sussell also testifies that his
system advertised Dr. Kantor's software, and made the software and user
manual publicly accessible. *Id*. at ¶ 18. According to Mr. Sussell, his
system had over 3,000 subscribers, in the 1993 timeframe, and the users had
keyword searching capability to retrieve Kantor. *Id*. at ¶¶ 6, 21.

Although we are cognizant that electronic documents generally are
not self-authenticating, it has been recognized that "[t]o authenticate
printouts from a website, the party proffering the evidence must produce
some statement or affidavit from someone with knowledge of the website
. . . for example a web master or someone else with personal knowledge
would be sufficient." *St. Luke's Cataract and Laser Institute v. Sanderson*,

2006 WL 1320242, *2 (M.D. Fla. 2006) (quoting *In re Homestore.com, Inc. Sec.Litig.*, 347 F. Supp. 2d 769, 782 (C.D.Cal. 2004)) (internal quotation marks omitted); *see also Market-Alerts Pty. Ltd. v. Bloomberg Finance L.P.*, 922 F. Supp. 2d 486, 493, n.12 (D. Del. 2013) (citing *Keystone Retaining Wall Sys., Inc. v. Basalite Concrete Prods., LLC*, 2011 WL 6436210, *9 n.9 (D. Minn. 2011)) (Documents generated by a website called the Wayback Machine have been accepted generally as evidence of prior art in the patent context); *U.S. v. Bansal*, 663 F.3d 634, 667-68 (3d. Cir. 2011) (concluding that the screenshot images from the Internet Archive were authenticated sufficiently under Federal Rule of Evidence 901(b)(1) by a witness with personal knowledge of its contents, verifying that the screenshot the party seeks to admit are true and accurate copies of Internet Archive's records).

Here, Mr. Sadofsky, who is a technology archivist and software historian, and currently is an archivist for the Internet Archive, testifies that he launched website textfiles.com and subdomain cd.textfiles.com to collect software, data files, and related materials from Bulletin Board Systems. Ex. 1081 ¶¶ 9-11. According to Mr. Sadofsky, textfiles.com and cd.textfiles.com are dedicated to preserving, archiving, and providing free access to unaltered historical software programs and information that initially were made available on the Bulletin Board Systems. *Id.* Mr. Sadofsky states that he previously archived the FWKCS Zip file (FWKCS122.ZIP) that contains Dr. Kantor's software and user manual to cd.textfiles.com from his own copy of the *Simtel MSDOS Archive*, October 1993 Edition, Walnut Creek CD-ROM. *Id.* at ¶ 14 (citing

Ex. 1052). Mr. Sadofsky also testifies that he personally verified the authenticity of Kantor—version 1.22, the version relied upon by EMC (Ex. 1004)—by comparing it with the "1993 archived" version and determined that Kantor is identical to the "1993 archived" version. Ex. 1081 ¶¶ 13-15. Mr. Sadofsky confirms that the source file of the "1993 archived" version has a timestamp of August 10, 1993, at 1:22 AM. *Id.* at ¶ 16; Ex. 1091 ¶¶ 10-11; Ex. 2014 ¶ 5. Mr. Sadofsky concludes that Kantor was publicly accessible prior to the critical date. Ex. 1081 ¶¶ 13, 16. PersonalWeb does not present sufficient or credible evidence to the contrary. Based on the evidence before us, we determine that Kantor has been authenticated sufficiently to warrant its admissibility under Federal Rules of Evidence 901(b)(1), (b)(3), and (b)(4).

With respect to Langer, Mr. Moore, who has personal knowledge of the operation of Usenet in 1991, testifies that Langer's header is consistent with the format of Usenet articles from the 1991 time frame, and the "Date:" field—indicating that Langer was posted on August 7, 1991, at approximately 10:51 PM GMT—would have generated automatically when the article was posted to Usenet. Ex. 1059 ¶ 16. Mr. Moore also testifies that he personally verified the authenticity of Langer by comparing it with an archived version obtained from Google Groups, which contains a compilation of Usenet articles going back to the 1980s and is recognized as a key archive of Usenet articles. *Id.* at ¶ 19. Accordingly, Langer has been authenticated sufficiently to warrant its admissibility under Federal Rules of Evidence 901(b)(1), (b)(3) and (b)(4).

We are not persuaded by PersonalWeb's argument that the download
date of "7/29/2003" in the lower, right-hand corner calls into question
whether Langer existed prior to the critical date. The mere fact that a
"downloaded" copy of Langer has a date subsequent to earliest effective
filing date is not sufficient to rebut EMC's supporting evidence that Langer
is what it claims to be—an article posted on Usenet newsgroups on August
7, 1991. *See, e.g.*, Ex. 1059 ¶¶ 11-17.

Moreover, we agree with EMC that Kantor and Langer also have been
authenticated as an "ancient document" under Federal Rule of Evidence
901(b)(8).[5] Opp. 6, 9. Each document is at least 20 years old and can be
found in a place where an authentic 20-year old document distributed
through a Bulletin Board System or Usenet would likely be. Ex. 1081
¶¶ 7-8; Ex. 1059 ¶ 19; *see also* Fed. R. Evid. 901(b)(8) 2012 Adv. Comm.
Note ("The familiar ancient document rule of the common law is extended
to include data stored electronically or by other similar means."ation).
Furthermore, testimony of Messrs. Sussell, Sadofsky, and Moore has
established sufficiently that the documents are in a condition that creates no
suspicion about their authenticity. Exs. 1053, 1081, 1059. Accordingly, we

---

[5] Fed. R. Evid. 901(b)(8). Evidence About Ancient Documents or Data
Compilations. For a document or data compilation, evidence that it:
    (A) is in a condition that creates no suspicion about its authenticity;
    (B) was in a place where, if authentic, it would likely be; and
    (C) is at least 20 years old when offered.

conclude that Kantor and Langer also have been authenticated sufficiently to warrant its admissibility under Federal Rule of Evidence 901(b)(8).

In addition, we are not persuaded by PersonalWeb's hearsay arguments. As EMC notes, a prior art document submitted as a "printed publication" under 35 U.S.C. § 102(b) is offered simply as evidence of what it described, not for proving the truth of the matters addressed in the document. *See Joy Techs., Inc. v. Manbeck*, 751 F. Supp. 225, 233 n.2 (D.D.C. 1990), *judgment aff'd*, 959 F.2d 226 (Fed. Cir. 1992); Fed. R. Evid. 801(c) 1997 Adv. Comm. Note ("If the significance of an offered statement lies solely in the fact that it was made, no issue is raised as to the truth of anything asserted, and the statement is not hearsay."). Therefore, neither Kantor nor Langer is hearsay under Federal Rule of Evidence 801(c).

We further agree with EMC that the posted and copyright dates set forth in Kantor and Langer are not a basis for excluding the documents, as testimony from Messrs. Sussell, Sadofsky, and Moore sufficiently establishes that Kantor and Langer existed prior to the critical date. Further, the computer-generated timestamp—August 10, 1993, at 1:22 AM—of the "1993 archived" version of Kantor (Ex. 1081 ¶¶ 14-15; Ex. 1091 ¶¶ 10-11; Ex. 2014 ¶ 5) also independently corroborates Kantor's existence as of August 10, 1993. *See, e.g.*, *U.S. v. Khorozian*, 333 F.3d 498, 506 (Fed. Cir. 2003) (concluding that an automatically generated time stamp on a fax was not a hearsay statement because it was not uttered by a person). Accordingly, we are not persuaded that PersonalWeb has presented a sufficient basis to exclude Kantor and Langer as impermissible hearsay.

For the foregoing reasons, we decline to exclude Kantor and Langer.

2. <u>Documents Corroborating Witnesses' Knowledge and Recollections</u>

PersonalWeb asserts that a number of documents submitted by EMC (Exs. 1050-1052, 1055-1058, 1061-1064, 1073, 1077, 1078, 1083-1085), and the declarations of Messrs. Sussell and Sadofsky (Exs. 1053, 1081, 1091) regarding those documents should be excluded, because the documents have not been authenticated properly and are inadmissible hearsay. PO Mot. 6-9. PersonalWeb argues that EMC "has not established that any of these documents existed prior to the critical date, and no witness has personal knowledge of their alleged existence prior to April 11, 1995." *Id*. at 7. PersonalWeb further maintains that the documents that are Exhibits 1056, 1057, 1077, and 1078 are irrelevant, prejudicial, and confusing, as they discuss a version of Kantor different than the version relied upon by EMC (version 1.22, Ex. 1004). *Id*. at 8-9.

EMC responds that its witnesses provided those documents to corroborate their independent knowledge and recollections. Opp. 10. EMC asserts that the documents have been authenticated under Federal Rules of Evidence 901-902, and fall within a hearsay exception under Federal Rules of Evidence 803-807. *Id*. at 10-12. We are persuaded by EMC's arguments.

As the movant, PersonalWeb has the burden of proof to establish that it is entitled to the requested relief. 37 C.F.R. § 42.20(c). As discussed previously, we disagree with PersonalWeb's argument that documents cannot be authenticated without direct testimony from the author or a witness who actually reviewed the documents prior to the critical date.

See Fed. R. Evid. 901(a). Significantly, PersonalWeb's motion does not explain sufficiently why each document should be excluded. For instance, PersonalWeb does not explain adequately why the declaration of Mr. Sussell (Ex. 1053 ¶¶ 3-4, 7, 10-11) is not sufficient to authenticate Exhibits 1055-1058, why the declarations of Mr. Sadofsky (Ex.1081 ¶¶ 3, 4, 6-8; Ex. 1091 ¶¶ 4-9) are not sufficient to authenticate Exhibits 1050-1052 and 1083-1085, or why the declaration of Mr. Moore (Ex 1059, ¶¶ 4-14) is not sufficient to authenticate Exhibits 1061-1064, 1073. *See* Fed. R. Evid. 901(b)(1).[6] Nor does PersonalWeb explain sufficiently why certain documents are not self-authenticated: (1) Exhibits 1055-1057, 1077-1078 – documents that have trade inscriptions; and (2) Exhibit 1052 – a photograph of the *Simtel MSDOS Archive*, October 1993 Edition, Walnut Creek CD-ROM, that has Simtel trade inscriptions. *See* Fed. R. Evid. 902(6)-(7).[7]

In its motion, PersonalWeb also fails to identify, specifically, the textual portions of the aforementioned exhibits that allegedly are being

---

[6] Fed. R. Evid. 901(b)(1). Testimony of a Witness with Knowledge.
    Testimony that an item is what it is claimed to be.

[7] Fed. R. Evid. 902. Evidence that Is Self-Authenticating
The following items of evidence are self-authenticating; they require no extrinsic evidence of authenticity in order to be admitted:
    . . . .
    (6) Newspapers and Periodicals. Printed material purporting to be a
    newspaper or periodical.
    (7) Trade Inscriptions and the Like. An inscription, sign, tag, or label
    purporting to have been affixed in the course of business and
    indicating origin, ownership, or control.

offered for the truth of the matter asserted, yet seeks to exclude the entirety of each exhibit. The burden should not be placed on the Board to sort through the entirety of each exhibit and determine which portion of the exhibit PersonalWeb believes to be hearsay. Rather, PersonalWeb should have identified, in its motion, the specific portions of the evidence and provided sufficient explanations as to why they constitute hearsay. Additionally, PersonalWeb does not explain adequately why the declarations of Messrs. Sussell, Sadofsky, and Moore do not provide the proper foundation and corroboration for the documents.

To the extent PersonalWeb relies upon the same arguments with respect to Kantor for excluding the documents, we have addressed those arguments above and determined that they are unavailing. We also agree with EMC that the documents concerning prior versions of Kantor are relevant, and not prejudicial or confusing as alleged by PersonalWeb, because such circumstantial evidence provides context and corroboration for the witnesses' independent knowledge and recollection.

Furthermore, we are not persuaded that the declarations of Messrs. Sussell, Sadofsky, and Moore (Exs. 1053, 1081, 1091, 1059) should be excluded. As we discussed above, and we elaborate below in the next section, Messrs. Sussell, Sadofsky, and Moore have sufficient personal knowledge and working experience to provide competent testimony to establish the publication and authentication of Kantor and Langer. The documents they cite serve to corroborate their independent knowledge and recollection.

For the foregoing reasons, PersonalWeb has not presented a sufficient basis to exclude Exhibits 1050-1052, 1055-1058, 1061-1064, 1073, 1077, 1078, 1083-1085, as well as the declarations of Messrs. Sussell, Sadofsky, and Moore (Exs. 1053, 1081, 1091, 1059), which include testimony concerning those exhibits.

3. Declarations of Messrs. Sussell, Sadofsky, and Moore

PersonalWeb argues that the declarations of Messrs. Sussell, Sadofsky, and Moore (Exs. 1053, 1081, 1091, 1059) should be excluded as hearsay under Federal Rule of Evidence 801 and inadmissible under Federal Rules of Evidence 802-807 for lack of foundation and personal knowledge, and Federal Rule of Evidence 702 as improper testimony, because the witnesses personally did not review Kantor (Ex. 1004), Simtel (Ex. 1052), and Langer (Ex. 1003) prior to the critical date. PO Mot. 9. PersonalWeb also argues that Messrs. Sussell, Sadofsky, and Moore "are not qualified experts" in the field. *Id*. at 10. PersonalWeb further alleges that Mr. Sadofsky's deposition (Ex. 2013, 30, 66) should be excluded, as it was responsive to a leading question (*id*. at 65-66) and non-responsive to another question (*id*. at 30-31). PO Mot. 10-11.

EMC responds that the testimony of Messrs. Sussell and Sadofsky should not be excluded, because their testimony is based on their own personal knowledge and recollection, and the documents they cite serve to corroborate their independent knowledge and recollection. Opp. 12-13. EMC further explains that the witnesses have described thoroughly the underlying facts, and, therefore, the testimony should be admitted as relevant

under Federal Rules of Evidence 401-402, supported by personal knowledge
and foundation under Federal Rule of Evidence 602, and proper opinion
testimony under Federal Rules of Evidence 701-703. We find that EMC's
contentions have merit.

PersonalWeb's arguments rest on the erroneous premise that EMC's
witnesses must have reviewed Kantor, Simtel, or Langer, personally *prior to
the critical date* in order to provide competent testimony regarding Kantor,
Simtel, or Langer. As discussed previously, it is well settled that it is not
necessary for the witnesses to have reviewed the reference personally *prior
to the critical date*, in order to establish publication. *See, e.g., Wyer*, 655
F.2d at 226. Although Messrs. Sussell, Sadofsky, and Moore are not experts
related to the claimed subject matter of the '539 patent, each witness
nevertheless has sufficient personal knowledge and working experience to
provide competent testimony. *See Hall*, 781 F.2d at 899. Mr. Sussell was
the co-owner and system operator of the Invention Factory Bulletin Board
System from 1983 to 1996. Ex. 1081 ¶ 3. Mr. Sussell's testimony is based
on his personal knowledge of the relevant facts related to the Invention
Factory Bulletin Board System and its association with Kantor. *Id*. ¶ 2.
Notably, Dr. Kantor specifically thanked Mr. Sussell in his user manual for
hosting Dr. Kantor's software FWKCS and for Mr. Sussell's role in its
development. Ex. 1004, 3 ("To Michael Sussell, sysop of The Invention
Factory (R), home board for the support of FWKCS, for bringing the
problem of duplicate files to my attention and for his help in testing . . . .").

Mr. Sadofsky is a technology archivist and software historian, and works for the Internet Archive, which provides the Wayback Machine service. Ex. 1081 ¶ 3. Mr. Sadofsky also directed an eight-episode documentary film regarding the Bulletin Board Systems. *Id*. at ¶ 4. Mr. Sadofsky's testimony is based on his personal knowledge of the relevant facts related to Kantor and the "1993 archived" version of Kantor. *Id*. at ¶ 2; Ex. 1091 ¶ 2. For example, Mr. Sadofsky personally verified the authenticity of Kantor by comparing it with the "1993 archived" version, and determined that Kantor is identical to the "1993 archived" version. Ex. 1081 ¶¶ 14, 15.

Similarly, Mr. Moore has personal knowledge and working experience with Usenet in 1991. Ex. 1059 ¶¶ 13-16. Mr. Moore's testimony is based on his personal knowledge of the relevant facts related to Usenet and its association with Langer. *Id*. at ¶ 10. Notably, Mr. Moore was intimately familiar with the operation of Usenet in the 1991-1992 timeframe, and he personally verified the authenticity of Langer by comparing it with an archived version obtained from Google Groups. *Id*. at ¶ 19.

Upon review of the evidence on the record, we also agree with EMC that Messrs. Sussell, Sadofsky, and Moore have disclosed sufficient underlying facts to support their testimony. For instance, the computer-generated timestamp—August 10, 1993, 1:22 AM—associated the "1993 archived" version of Kantor corroborates the testimony of Messrs. Sussell and Sadofsky regarding Kantor's existence as of August 10, 1993. Ex. 1081 ¶¶ 14-15; Ex.1091 ¶¶ 10-11; Ex. 2014 ¶ 5.

As to Mr. Sadofsky's deposition, PersonalWeb does not explain sufficiently why that testimony should be excluded. PO Mot. 11. Moreover, Mr. Sadofsky's deposition (Ex. 2013, 30, 66) is consistent with his direct testimony (Ex. 1081 ¶¶ 14-16), and, therefore, it would not prejudice PersonalWeb even if such evidence is not excluded.

For the foregoing reasons, PersonalWeb has not presented a sufficient basis to exclude any portion of the declarations of Messrs. Sussell, Sadofsky, and Moore (Exs. 1053, 1081, 1091, 1059) and Mr. Sadofsky's deposition (Ex. 2013, 30, 66).

4. Clark's Rebuttal Declaration

PersonalWeb asserts that Dr. Clark's rebuttal declaration (Ex. 1092) should be excluded, because it is irrelevant, prejudicial, and confusing, as well as beyond the scope of this proceeding. PO Mot. 11-15. In support of its assertion, PersonalWeb advances several arguments. *Id*.

First, PersonalWeb argues that Dr. Clark's rebuttal declaration cites to references that do not serve as the basis of a ground of unpatentability instituted in this proceeding. *Id*. at 11-12. EMC counters that Dr. Clark's statements referencing those references were offered in response to PersonalWeb's argument that one with ordinary skill in the art would not have modified Kantor or Woodhill. Opp. 13 (citing PO Resp. 15-22, 51-54; Ex. 2020 ¶¶ 42-48, 94-98). According to EMC, those statements are relevant to the instituted grounds of unpatentability and confirm that the use of hash-based identifiers to identify files was well known in the art at the time of invention. *Id*. We agree with EMC that Dr. Clark's statements are

proper rebuttal evidence submitted in response to PersonalWeb's arguments.

Those references were cited merely to show the knowledge level of a person

with ordinary skill in the art. *See Randall Mfg. v. Rea*, 733 F.3d 1355, 1362

(Fed. Cir. 2013) (When considering whether a claimed invention would have

been obvious, "the knowledge of [an ordinarily skilled] artisan is part of the

store of public knowledge that must be consulted."). Such evidence does not

change the combination that formed the basis of the grounds of

unpatentability based on obviousness instituted in this proceeding. *Id*.; *see*

also In re Donohue, 766 F.2d 531, 534 (Fed. Cir. 1985). Accordingly, we

are not persuaded that PersonalWeb has presented a sufficient basis to

exclude Dr. Clark's rebuttal declaration.

Second, PersonalWeb contends that the "capable," "can," and "may"

statements in Dr. Clark's rebuttal declaration should be excluded, because

those statements are irrelevant, prejudicial, confusing, lacking foundation,

and beyond the scope of this proceeding. PO Mot. 12. PersonalWeb further

submits that Dr. Clark's rebuttal declaration includes new obviousness

allegations not presented previously with the petition. *Id*. at 12-13. In

response, EMC contends that the statements in Dr. Clark's rebuttal

declaration were offered in response to PersonalWeb's arguments. Opp.

13-15 (citing *e.g.*, PO Resp. 9-10; Ex. 2020 ¶ 31-32). Having reviewed

PersonalWeb's patent owner response and Dr. Clark's rebuttal declaration,

we determine that Dr. Clark's testimony is reasonable rebuttal evidence in

light of PersonalWeb's arguments. Furthermore, PersonalWeb's arguments

concerning Dr. Clark's statements affect the weight to be given by us to

Dr. Clark's testimony in deciding whether the instituted grounds of unpatentability render the challenged claimed unpatentable. When weighing evidence, we are capable of determining whether the prior art references anticipate or render obvious the challenged claims without being confused, misled, or prejudiced by Dr. Clark's testimony. Thus, we are not persuaded that PersonalWeb has presented a sufficient basis to exclude any portions of Dr. Clark's rebuttal declaration.

Finally, PersonalWeb contends that Dr. Clark's rebuttal declaration contradicts his prior deposition. PO Mot. 13-15. We are not persuaded by PersonalWeb's arguments. Rather, we agree with EMC that Dr. Clark's rebuttal testimony is consistent with his earlier testimony. Opp. 15. For instance, Dr. Clark's rebuttal testimony that "zipfiles are not *always* compressed," and the inner files of a zip file may be *uncompressed* (Ex. 1092 ¶¶ 9-11), is consistent with his earlier testimony that the inner files of a zip file are compressed *typically* (Ex. 2016, 55, 59, 66-67). Moreover, Dr. Clark's testimony is reasonable rebuttal evidence in light of the evidence submitted by PersonalWeb. Dr. Clark merely points out in his rebuttal declaration that PersonalWeb's evidence also shows that zip files are not *always* compressed. Ex. 1092 ¶ 9 (citing Ex. 2004, 3 (the zip file format defines seven compression methods which include "Compression method 0" that does not compress the file); Ex. 1088, 262 (Dr. Dewar agrees that "the zipfile standard allows for uncompressed files.")).

In addition, we agree with EMC that Dr. Clark's testimony does not conflict with EMC's position advanced in its petition that the inner files in

Kantor constitute the relevant portion of the zip file for determining segment identifiers. Opp. 15 (citing Pet. 44; Ex. 1009 ¶ 35). We do not discern that Dr. Clark's answer to a question related to "a sequence of *people*" (Ex. 2016, 94-98) contradicts with Dr. Clark's rebuttal testimony on "a sequence of *bits*" of a data item (Ex. 1092 ¶ 28). Dr. Clark in the prior deposition also testified that there are examples of sequences with intervening gaps including Fibonacci sequences, random sequences, odd sequences, and even sequences. Opp. 15 (citing Ex. 2016, 191-193). Accordingly, we are not persuaded that PersonalWeb has presented a sufficient basis to exclude the alleged inconsistent statements in Dr. Clark's rebuttal declaration.

For the foregoing reasons, we decline to exclude Dr. Clark's rebuttal declaration (Ex. 1092).

III. CONCLUSION

EMC has met its burden of proof, by a preponderance of the evidence, in showing that claims 10, 21, and 34 the '539 patent are unpatentable based on the following grounds of unpatentability:

Claim	Basis	References
10 and 21	§ 102(b)	Langer
34	§ 103(a)	Langer and Woodhill
10 and 21	§ 103(a)	Kantor
34	§ 103(a)	Kantor and Langer
10 and 21	§ 103(a)	Woodhill and Fischer

IV. ORDER

In consideration of the foregoing, it is

ORDERED that claims 10, 21, and 34 of the '539 patent are held *unpatentable*;

FURTHER ORDERED that EMC's Motion to Exclude Evidence is *dismissed*;

FURTHER ORDERED that PersonalWeb's Motion to Exclude Evidence is *denied*; and

FURTHER ORDERED that, because this is a final written decision, parties to the proceeding seeking judicial review of the decision must comply with the notice and service requirements of 37 C.F.R. § 90.2.

PETITIONER:

Peter M. Dichiara, Esq.
David L. Cavanaugh, Esq.
WILMER CUTLER PICKERING HALE & DORR LLP
peter.dichiara@wilmerhale.com
david.cavanaugh@wilmerhale.com

PATENT OWNER:

Joseph A. Rhoa, Esq.
Updeep. S. Gill, Esq.
NIXON & VANDERHYE P.C.
jar@nixonvan.com
usg@nixonvan.com